International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

Flight Ticket Price Prediction Using Machine Learning

Prof.Ghadge S.S[1], Kamble Priya [2], Thorat Komal [3]

Professor, Computer Department, SPCOETSomeshwarnagar, Baramati, India

Student, Computer Department, SPCOETSomeshwarnagar,Baramati, India

Student, Computer Department, SPCOETSomeshwarnagar,Baramati, India

---

ABSTRACT

As technology is growing rapidly travelling has become a important part our life. Flight ticket price is very changeable data that is updating on daily basis. Based on various parameters it changes. we cannot say that prices of ticket remains same or not. It may change within week, month or some days. A lot of factors affect on prices of airline ticket like departure date, arrival date, number of stoppages, airline that you want and so on. The proposed framework uses the machine learning algorithm to model the quarterly moderate ticket price based on the different sources and destinations. It help customers to predict future flight prices and plan their journey accordingly

Key Words: Machine Learning, Random Forest Algorithm, Pycharm, Anaconda

INTRODUCTION

Nowadays airline corporations are using complex strategies for flight ticket fare calculations. This highly complicated method makes the flight ticket fare difficult to guess for the customers, since the fare changes dynamically. A Flight price prediction system that predicts the price of a flight for a particular date based on various parameters like Source, Destination, Stoppages & Airline. Our project Improved Flight Price Prediction System which resolves this problem and provides a facility where people will be able to predict the flight ticket price before purchasing the ticket

MOTIVATION

Motivation is to help people to predict the flight prices with comparison of today to another any day because of this customer can be book their tickets of Flight according to their comfortably, according to their affordability, and to save their time and money so that user can book the flight ticket according to need.

OBJECTIVE

The prime objective of this project is to use machine learning techniques to model the behaviour of flight ticket prices over time and predict the price of the flight-ticket .The objective of this project is to study how Airline ticket prices change over time based on different parameters.

PROBLEM STATEMENT

Flight ticket prices are the most changeable data that changes every day depending on various parameters like departure date, arrival date, number of stoppages and airline. We always heard travelers saying that flight ticket prices are so unstable. As data scientists, we are going to prove that by giving the right data, anything can be predicted. Here you will be provided with prices of flight tickets for various airlines between the months. As a result, many people will save time and money by having a basic understanding of flights before making travel arrangements. A predictive model will be created in the proposed system by application of machine learning algorithms to collect historical data.

LITERATURE SURVEY

A literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic. Literature reviews use secondary sources, and do not report new or original experimental work.

1.Paper Name:A Framework for Airfare Price Prediction: A Machine Learning Approach

Author: Tianyi Wang, SamiraPouyanfar, Haiman Tian, YudongTao†, Miguel Alonso Jr., Steven Luis and Shu Ching Chen School of Computing and Information Sciences Florida International University, Miami, Florida 33199

To the best of our knowledge, most of previous studies on airfare price prediction using the DB1B dataset have focused on conventional statistical approaches, which have their own limitations of problem estimations and assumptions

2.Paper Name:A Bayesian Predictor of Airline Class Seats Based on Multinomial Event Model

Author: Yudong Tan and Huimin Zhou Ctrip.com Shanghai, China

In this paper we investigated the airline ticket prices on the view of class seats' inventory reservations instead of building a direct price predictor like most research do The model prediction results can be embedded into search engine layer to give effective buy-or wait suggestion to customers, thus the predictor devotes itself into an instance of big data application.

3.Paper Name: An Airfare Prediction Model for Developing Markets

Author: Viet Hoang Vu, Quang Tran Minh Faculty of Computer Science and Engineering Ho Chi Minh City University of Technology, VNU - HCM Ho Chi Minh City, Viet Nam

In this paper, we have evaluated several conventional machine learning algorithms on our airfare dataset to build an interpretable prediction model that can predict the trend of airfare in an emerging deviation market (Vietnam) to help customers to decide when they should purchase airfare to get as most saving as they can. We used data collected from the websites that sell airfare online using our own tool.

4.Paper Name: Predicting Airfare Price Using Machine Learning Techniques: A Case Study for Turkish Touristic Cities

Author: Bilgisayar M¨uhendisligˇı B¨ol¨um¨u, Boˇgazi¸ci Universitesi, Is- tanbul, Turkiy

Airline ticket price is influenced by several elements, such as flight distance, purchasing time, number of transfers, etc. Furthermore, every carrier has its own proprietary rules and techniques to determine the ticket price accordingly. With recent improvements in Machine Learning (ML)

5.Paper Name: Liposomal drug delivery systems: From concept to clinical applications

Author: T.M. Allen and P.R. Torchilin

In this paper, we propose a new model that can help the buyer to predict the price trends without official information from the airlines. Our findings demonstrated that the proposed model can predict the trends as well as actual airfare's changes up to the departure dates using public airfare data available online despite the missing of many key features like the number of unsold seats on flights. We also identified the features that have the strongest impacts on the airfare change

PROPOSED SYSTEM

Predicting the price of an airline ticket is a very difficult task because many factors depend on the price of an airline ticket. Many researchers used various machine learning algorithms to get a model with higher prediction accuracy from the ticket price.

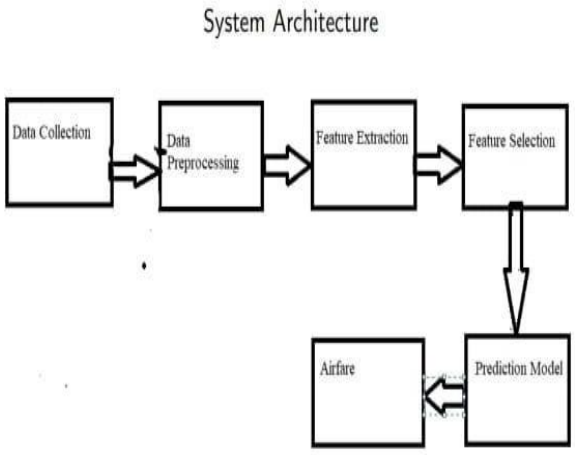

Figure. Proposed System

METHODOLOGY

The first step is the collection of data, where the historical flight data is collected for a ticket price prediction model. Then we cleansed the dataset during the exploratory data analysis process by eliminating duplicate and null values. The accuracy of the model would suffer if these values weren't eliminated. The next step is data pre-processing, where we noticed that string format was used to store the majority of the data. Every feature's data is retrieved, such as the day and month from the trip's date in integer format and the hours and minutes from the departure time. Source and destination features had to be transformed to values because they were of the categorical. The feature selection step is involved in selecting the important properties that correlate more with price. There are some features such as additional information and route that are unnecessary features that can affect accuracy model and therefore need to be removed before obtaining Our model ready for prediction.

ALGORITHM USED

Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees.[citation needed] However, data characteristic scan affect their performance. What is Random Forest used for? Random forest is used on the job by data scientists in many industries including banking, stock trading, medicine, and e-commerce. It's used to predict the things which help these industries run efficiently.

(A) Feature Selection

Phase one (Feature Selection) - throughout this part the foremost informative options of a flight verify the costs of the air ticket area unit set. This part is extremely vital since it defines the matter underneath determination. For every flight, the subsequent options were considered: F1: Feature one - time of departure. F2: Feature two-point. F3: Feature three - ranges of free baggage (0, 1, or 2). F4: Feature four - days left till departure.F5: Feature five - ranges of intermediate stops. F6: Feature half-dozen - vacation days (yes or no). F7: Feature a seven-night long flight (yes or no). F8: Feature eight - days of the week.
It is a price to notice that the influence of some essential features from the higher than list is examined through a "one leave-out "rule. We tend to additionally wish to clarify that the feature F4indicates the number of days between the price ticket purchase and the day of the flight. Phase two (Data Collection) - during this study, our interest is focused on the prediction of one transportation value while not return.

(B) Random Forest Classifier

Random Forest could be a versatile, easy-to-use machine learning formula that produces, even while not hyper- parameter standardisation, a superb result most of the time. It is also one of the most widely used algorithms, due to its simplicity and versatility. Random Forest is a supervised learning algorithm. The "forest "is constructive, a collection of tropical trees, usually trained in the "enclosing" method. A common idea of how to pack a bag is theta combination of learning models enhances the overall effect. The informal forest forms many decision-making trees and combines them together to obtain the most accurate and stable prediction. The main unit of random forest sections is the decision tree. Decision tree is a hierarchical structure constructed using the elements (or independent variants) of a data set. Each node of the decision tree is divided into a scale associated with a small 27 set of features. Random Forest could be an assortment of call trees related to a collection of bootstrap samples generated from an explicit information set. The nodes are categorized based on the entropy (or Gini index) of the selected subset of features. The subset sets created from the original data set, using bootstrapping, are the same size as the original data set

PERFORMAMCE METRICS

The accuracy of machine learning models trained by various algorithms will be compared using performance metrics, which are statistical models. Regression metrics will be implemented for error measurement functions from each model using the sklearn. Metrics module. The following metrics will be examined to determine each model's error rate:

MAE (Mean Absolute Error)

The mean of the absolute difference between the expected and actual numbers is effectively added to determine the mean absolute error.

MAE = 1/n [∑ (y-ý)]

The expected output values are y' and the actual output values are y.

There are n total data points.

Your model will perform better the lower the MAE number is.

MSE (mean square error)

The root mean square error exponentiates the difference of the true a predicted output values before summing them Instead using an absolute value.

y=actual output values

ý=predicted output values

n = Total number of data points

MSE penalizes large errors when we square the errors. Less the MSE value, the better the model performance

RMSE (root mean square error)

RMSE is measured by taking the square root of the mean squared difference between forecast and actual value.

RMSE = √1/n [∑ (y-ý) 2]

The expected output values are y' and the actual output values are y.

There are n total data points. The higher the performance of a model, the more RMSE is bigger than MAE and smaller than RMSE value comparing different mode

R2 (Coefficient of determination)

It will help you understand how well the independent variable modified with a deviation in your model.

R2 = 1 −∑(ý-ȳ)2 / ∑(y-ȳ)2

The R-squared value lies between 0 and 1. The closer its value is for one, the better your model is compared to others model values

MATHEMATICAL MODEL

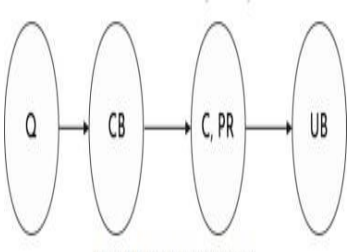

Figure 1: Mathematical Model.

- Q:Input CSV File
- CB:Preprocessing
- C:Feature Extraction
- PR :Classification
- UB: Output

1) **The value of Random Forest tree determine by following way:**

$$fi_i = \frac{\sum j: \text{node } j \text{ splits on feature } i^{n_{ij}}}{\sum k \in \text{all nodes } n_{i_k}}$$

2) **The value of function determine and divided by total number of tree:**

$$RF\, fi_i = \frac{\sum i \in \text{alltrees}^{normf_{ij}}}{T}$$

Success Conditions: Flight Ticket Prediction successfully.

Failure Conditions: Flight Ticket Prediction not successfully.

RESULT

CONCLUSION

In this study, a machine learning framework was developed to predict the moderate quarterly airfare price at the market segment level. It will help the customer with the flight ticket booking. Travelers can save money if they choose to buy a ticket when its price is the lowest.

FUTURE SCOPE

In the future, our plan can be extended to include air ticket transaction information, which can provide more detail about a specific itinerary, such as time and date of departure and arrival, seat location, covered ancillary products, etc. Thus, event information will also be collected from various sources, which include social platforms and news agencies, to complement our prediction model. Additionally, we will investigate other advanced ML models.

REFERENCE

[1] Vijayarangan Natarajan and Shubham Sinha, "A Novel Approach: Airline Delay Prediction Using Machine Learning",IEEE 2018.

[2]T ianyi Wang, Samira Pouyanfar, Haiman Tian, Miguel Alonso Jr., Steven Luis and Shu-Ching Chen "A Framework for Airfare Price Prediction: A Machine Learning Approach", IEEE 2020.

[3] Jiage Huo and K.L. Keung , "The Prediction of Flight Delay: Big Data-driven Machine Learning Approach" , IEEE 2020.

[4] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola and V. Vapnik, ‖Support vector regression machines,‖ Advances in neural information processing systems, vol. 9, pp. 155-161, 1997.

[5] S. Haykin, Neural Networks – A Comprehensive Foundation. Prentice Hall, 2nd Edition, 1999.

[6] S.B. Kotsiantis, —Decision trees: a recent overview,‖ Artificial Intelligence Review, vol. 39, no. 4, pp. 261-283, 2013.

[7] [4] Viet Hoang Vu, Quang Tran Minh and Phu H. Phung,‖An Airfare Prediction Model for Developing Markets‖, IEEE paper 2018.

[8] T. Janssen, —A linear quantile mixed regression model for prediction of airline ticket prices,‖ Bachelor Thesis, Radboud University, 2014.

[9] 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.

[10] Supriya Rajankar and Neha Sakharkar, —A Survey on Flight Pricing Prediction using MachineLearning‖, International Journal of 56 "Flight Ticket Price Prediction Using Machine Learning" Engineering Research and Technology, vol 8, issue 6, June 2019. Diamantaras,‖Airfare Prices Prediction Using Machine Learning Techniques‖, EUSIPCO 2017..

[11]J. Stavins, "Price discrimination in the airline market: The effect of market concentration," Review of Economics and Statistics, vol. 83, no. 1, pp. 200–202, 2001

[12] B. Mantin and B. Koo, "Dynamic price dispersion in airline markets," Transportation Research Part E: Logistics and Transportation Review, vol. 45, no. 6, pp. 1020–1029, 2009.

[13] P. Malighetti, S. Paleari, and R. Redondi, "Has ryanair's pricing strategy changed over time? an empirical analysis of its 2006–2007 flights," Tourism Management, vol. 31, no. 1, pp. 36–44, 2010.

[14] T. H. Oum, A. Zhang, and Y. Zhang, "Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets," Journal of Transport Economics and Policy, vol. 7, no. 2, pp. 171–192, 1993

[15]Ingold A, Huyton J R.:Yield management and the airline industry. Yield Management: Strategies for the service industries, 2000: 179-197 (2000)