International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2020): 6.733

Implementation of Flight Fare Prediction Web App Project with Deployment

Aryan Madaan[1], Dr. S. Gnanavel[2]

[1]Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur-603203, Chengalpatu District, Tamil Nadu, India
E-mail: am8123[at]srmist.edu.in

[2]Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur-603203, Chengalpatu District, Tamil Nadu, India
E-mail: gnanaves1[at]srmist.edu.in

Abstract: ***Traveling via flights has become an integral part of today's lifestyle, as more and more people choose faster travel options. Airfare prices increase or decrease every day here and there depending on various factors such as flight timings, destination, duration of flights, various occasions such as holidays or holiday season. 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 collected historical data. This system will give people an idea of the trends the prices follow and also provide the predicted value of the price they can check before booking flights to save money. This kind system or service can be provided to customers through flight booking companies to help customer book tickets.***

Keywords: Machine learning; Prediction model; Feature selection, Airfare price; Pricing Models; Random Forest

1.Introduction

The ticketing system is to buy the ticket many days before the flight takes off to avoid the effects of the most extreme fees. Air routes usually do not agree with this procedure. Airlines can reduce costs at times when they need to build a market and when tickets are less affordable. They can maximize costs. So the price may depend on various factors. In order to predict the costs, this business uses AI to show the ways of the tickets after some time. All groups are free to adjust the price of their tickets at any moment. An explorer can save cash by booking the lowest cost flight. People who have traveled frequently by plane are aware of the fluctuations in prices.

Airlines use comprehensive Revenue Management principles to implement distinctive rating systems. As a result, the evaluation system changes the fee depending on the time, season and holidays to change the header or footer on the following pages. The ultimate goal of airways is to make a profit while the customer seeks the minimum rate. Customers usually try to buy a ticket well in advance of the departure date to avoid the increase in ticket prices as the date approaches. But in reality it is not. A customer may end up paying more than they should for the same location.

This project aims to predict flight prices for different flights using the machine learning model. The user receives the expected values, and using these as a guide, they can choose whether to purchase tickets.

In the current scenario, airlines are trying to manipulate ticket prices to maximize their profits. Many people frequently take flights, so they are aware of the optimum times to get affordable tickets. But there are also many people who have no experience while booking tickets and they end up falling into the trap of discounts from the companies where they actually end up spending more than they should have. By providing clients with the information they need to order tickets at the proper moment, the suggested system can help them save millions of rupees.

The proposed problem statement is “Flight price prediction System".

2.Literature Survey

Proposed Study [1] Airfare Price Prediction Using Machine learning techniques, for research paper data set consisting of 1814 Aegean Airlines data flights collected and used to train a machine learning model. Different a number of features were used to train different models demonstrate how feature selection can change accuracy Model.

In a case study [2] by William Groves, an agent is introduced which is able to optimize the timing of the purchase on behalf customers. To create a model, partial least squares regression is employed.

In [3] study, the desired model is implemented utilising the San Francisco-New York course's linear quantile mixed regression methodology, where tickets are listed every day through an online website. Two functions like the number of days for departure and whether the departure is on a weekend or a weekday are considered model development.

In proposed Study [4] we learn about different flight trends, the best time to buy a ticket. We have also successfully debunked some of the typical myths and misconceptions related to the airline industry and backed them up with data and analysis.

In a case study [5] by Tianya Wang, several features were extracted from the datasets and combined with the data to model the segments of the air transport market. With the help of feature selection techniques, our proposed model is able to predict the quarterly average ticket price.

In research done [6] by Vinod Kimbhaune, proper implementation of project has resulted in saving inexperienced people money by providing them with information regarding the trends that flight prices are following, as well as providing them with a predicted price value that they use to decide whether to book a ticket now or later.

In survey [7] by Neel Bhosale, Machine learning algorithms are applied to the data set to predict the dynamic price of flights. This gives the predicted values of the flight fare so that you can get the ticket at the minimum price.

Jaywrat Singh Champawat [8] proposed a framework to find a machine learning model that provides higher accuracy in predicting the price of Indian flights. Working with different models, it was found that the Random Forest algorithm showed the highest accuracy in predicting the output.

In [9] research, Fare prediction for civil aviation remains relatively imprecise and unreliable. A prediction method based on MADA is proposed to solve this problem. Judging from the experimental results, the MADA-based method can provide more accurate prediction results than traditional methods for civil aviation ticket prices.

Proposed Study [10] Square measure machine learning algorithms to predict the exact price of air tickets and provides the exact value of the air ticket price in the limited and highest value. Model accuracy is also predicted by the R-squared value.

In a case study [11] by QiqiRen, it focuses on aspects that are visible on the consumer side and only predicts a binary class of whether the price will increase or not, which is basically whether we should buy now or wait.

In the research done [12] by Juhar Ahmed Abdella, two main areas of research are discussed-prediction models that are designed to save money for the customer and those that are designed to increase airline revenue. The strengths and weaknesses of the existing work were discussed.

3.Proposed Work

Forecasting the price of an airline ticket is a very challenging task because many factors depend on the price of an airline ticket. Many researchers used various machine learning algorithms to obtain a model with higher prediction accuracy from the ticket price. Researchers use various regression models such as support vector machines (SVMs), Linear regression (LR), decision tree, random forests, etc. to predict the exact price of a flight.

After further reading, it was found that models are divided into two types-one that predicts minimum ticket price and one that helps generate maximum returns which can be termed as customer-side models and airline-side models. In addition to these, other researches were also carried out, such as researching the various factors that lead to changes in ticket prices and demand changes its price. These researches found that customers who travel for entertainment are more sensitive to ticket prices rather than customers traveling for business purposes. Date of reservation and date of travel is also being looked at by many researchers as influencing price increases. Studies are also being conducted on effects of delays on fares.

This article will use Python3 to implement machine learning algorithms to create a model that will make predictions with high precision. Various python libraries are imported to perform these actions.

There are various steps involved in building an ML model, starting with importing a dataset and cleaning the data. All null values and duplicate values are removed from the dataset. Then the data is encoded by conversion of some variables into a certain format. Converts categorical data to numeric data.

After the dataset is processed, feature selection is performed. Properties or variables that are not so important are removed from the dataset. Exploratory data analysis is performed to provide insight and identify important features using the Extra Tress Regressor. Feature Engineering is performed to reduce computational cost and sometimes to improve accuracy. This is done using a correlation matrix. Then the data is split into train and test data, where the train data is used to train the models. The test data is used to check the accuracy of the models. Then it decides on the optimal features and parameters performing hypertuning. After all models are trained, their accuracy is checked using their R-squared value.

4.Implementation

We implemented machine learning lifecycle for this project to create a basic web application that will predict flight prices using a machine learning algorithm with historical flight data using python libraries such as Pandas, NumPy, Matplotlib, seaborn and sklearn. Figure shows the steps that we were based on the life cycle:

The first step is the selection of data, where the historical flight data is collected for a price prediction model. Our dataset contains more than 10, 000 flight-related data

records and its prices. Source, destination, departure date and time, number of stops, arrival time, costs, and more are just a few of the dataset's functions.

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 preprocessing, 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 kind. For this One, categorical values are transformed into model-identifiable values using hot-coding and label coding approaches.
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.

The following phase involves employing a machine algorithm and model generation after features that are more closely related to price have been chosen. Since our dataset consists of labeled data, we will also use supervised machine learning algorithms under supervision we will use regression algorithms like ours the dataset contains continuous values in the functions. To explain the link between dependent and independent variables, regression models are used. We will utilise the following machine learning algorithms in our project:

Linear regression

We will use multiple linear regressions, which estimates the relationship between two or more independent variables and one dependent variable. In simple linear regression, there is only one independent and dependent function. However, our dataset contains many independent functions on which the price may depend.

The following depicts the multiple linear regression model:

$$\mathbf{Y = \beta 0x1 + \ldots. + \beta nxn + \mathcal{E}}$$

Y = the dependent variable's anticipated value

Independent variables = Xn

When all other parameters are zero, n = coefficients of independent variables equals the y-intercept.

Decision tree

Regression and classification trees are the two main forms of decision trees, where regression is used for continuous data and classification is used for categorical values. Decision the tree chooses an independent variable from the dataset as decision nodes.

The entire data file is divided up into many subsections, and when test data is fed into the model, the result is determined by determining which subsection the data point belongs to. The decision tree's output will be the average value of all the data points in the subsection, depending on which subsection the data point belongs to.

Random forest

In the Random Forest ensemble learning technique, the training model employs a number of different learning algorithms, and the separate outputs are then combined to produce the final anticipated outcome. The random forest belongs to the bagging category of ensemble learning, where a random number of elements and records are chosen and given to the model group. In essence, decision trees are used as a group of models in random forest. The average value of the anticipated values, if they are thought to be the output of the random forest model, can be calculated from the predictions produced by decision trees.

Performance 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.

$$\mathbf{MAE = 1/n \left[\sum (y\text{-}\acute{y})\right]}$$

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.

$$\mathbf{MSE = 1/n \left[\sum (y\text{-}\acute{y})\ 2\right]}$$

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 models.

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.

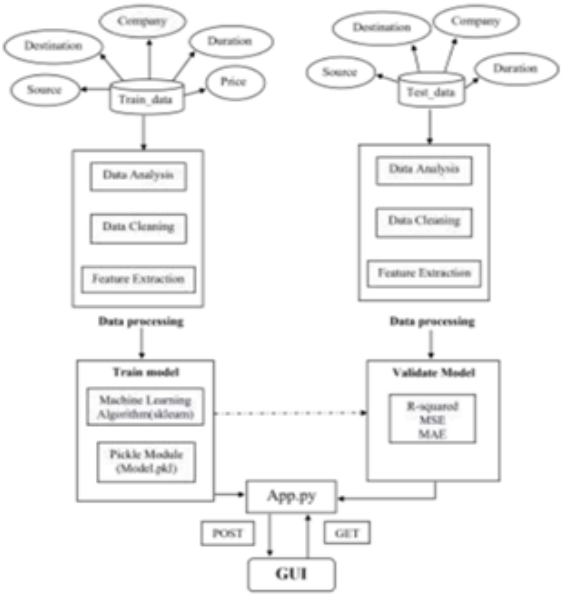


Figure: System Architecture Diagram

We include the last three steps of the life cycle model for deploying a trained machine learning model. Therefore, after obtaining the model with the best accuracy, we save this model to a file using the pickle module. The application will be built using the Flask Framework where API endpoints such as GET and POST will be created perform operations related to loading and displaying data on front-end applications.

The front-end of the application will be created using bootstrap framework where the user will have functions entering your flight details. This information will be transmitted to the backend service, where the model will forecast the result based on the input. The front-end receives the expected value and displays it.

5.Experimental Results and Conclusion

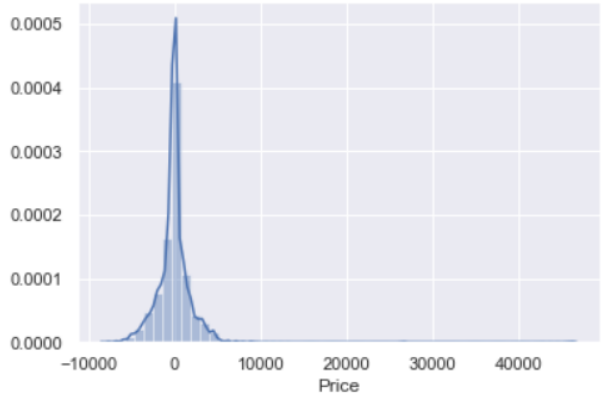
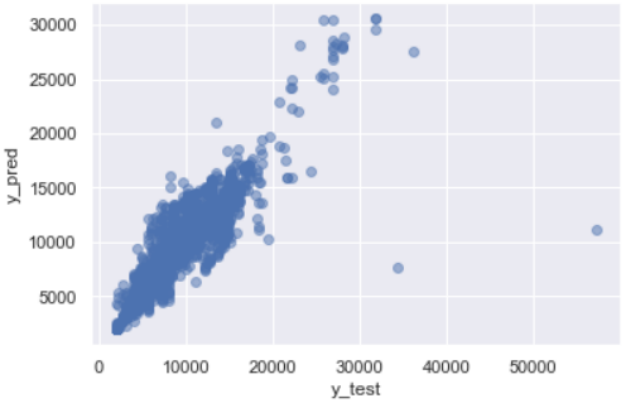


Figure: Graphical Results for Random Forest



As the graph is forming a Gaussian distribution, this means that our results are good.

This paper proposed to find a machine learning model that provides higher accuracy in fare price prediction of flights. Working with different models, it was found that the Random Forest algorithm showed highest output prediction accuracy. The paper provides better results than previously observed models and aims to improve in the future.

Proper implementation of this project can lead to saving inexperienced people's money by providing them information related to the trends that air fares follow and also give them the predicted value of the price they use to decide whether to book your flight now or later. In conclusion, this service can be implemented with good precision forecast. Because the predicted value is not completely accurate, there is a lot of room for improvement in this kind of service.

6.Future Scope

More routes can be added and the same analysis can be extended to major airports and travel routes in India. More of data points and historical data should be taken into

account for analysis. This will train the model better and provide better accuracy and more savings.

Additional rules may be added to rule-based learning based on our understanding of the industry, including offer periods provided by airlines. Development of a more user-friendly interface for different routes giving users more flexibility.

Currently, there are many fields where prediction services are used such as stock price prediction tools used by stockbrokers and services like Zestimate that provide estimated value of house prices. That's why there exists the demand for services like this in the aviation industry which can assist customers in booking tickets. There are many of them that examine the work that has been done on it using various techniques and further research is needed to improve prediction accuracy using different algorithms. Results can be more accurately obtained by using data that is more accurate and has better features.

References

[1] Joshi, Achyut & Sikaria, Himanshu & Devireddy, Tarun. (2017). Predicting Flight Prices in India.
[2] Wang, Tianyi & Pouyanfar, Samira & Tian, Haiman & Tao, Yudong & Alonso, Miguel & Luis, Steven & Chen, Shu-Ching. (2019). A Framework for Airfare Price Prediction: A Machine Learning Approach.200-207.10.1109/IRI.2019.00041.
[3] Supriya Rajankar, Neha Sakharkar, Omprakash Rajankar, “Predicting The Price Of A Flight Ticket With The Use Of Machine Learning Algorithms”, INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 8, ISSUE 12, ISSN 2277-8616, DECEMBER 2019.
[4] Neel Bhosale, Pranav Gole, Hrutuja Handore, Priti Lakde, Gajanan Arsalwad, “Flight Fare Prediction Using Machine Learning”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653, Volume 10, Issue V, May 2022.
[5] Jaywrat Singh Champawat, Udhhav Arora, Dr. K. Vijaya, “INDIAN FLIGHT FARE PREDICTION: A PROPOSAL”, International Journal of Advanced Technology in Engineering and Science, Vol. No.09, Issue No.03, ISSN 2348-7550, March 2021.
[6] Zhichao Zhao, Jinguo You, Guoyu Gan, Xiaowu Li, Jiaman Ding, “Civil airline fare prediction with a multi-attribute dual-stage attention mechanism”, Appl Intell **52**, 5047-5062 (2022), 03 August 2021.
[7] Pavithra Maria K, Anitha K L, “Flight Price Prediction for Users by Machine Learning Techniques”, International Advanced Research Journal in Science, Engineering and Technology, Vol.8, Issue 3, DOI: 10.17148/IARJSET.2021.8321, March 2021.
[8] Juhar Ahmed Abdella, NM Zaki, Khaled Shuaib, Fahad Khan, “Airline ticket price and demand prediction: A survey”, Journal of King Saud University-Computer and Information Sciences, Volume 33, Issue 4, May 2021.
[9] Rutuja Konde, Rutuja Somvanshi, Pratiksha Khaire, Prachi Zende, Kamlesh Patil, “Airfare Price Prediction System”, International Research Journal of Engineering and Technology (IRJET), Volume: 09 Issue: 05, May 2022.
[10] Vivekanand P. Thakare, Ankita Sanjay Murraya, Roshani Bandu Gawade, Mrunali Mukundrao Sawarkar, Trupti Khemraj Shende, Ujjwala Kamlesh Badole, “Flight Ticket Price Predictor Using Python”, Advancement and Research in Instrumentation Engineering Volume 5 Issue 1 e-ISSN: 2582-4341.
[11] Kunal Khandelwal, Atharva Sawarkar, Dr. Swati Hira, “A Novel Approach for Fare Prediction Using Machine Learning Techniques”, International Journal of Next-Generation Computing, VOLUME 12, SPECIAL ISSUE 5, Online ISSN: 0976-5034, NOVEMBER 2021.
[12] Prof. Ms. Archana Dirgule, Shubham Agarwal, Ram Agrawal, Neha Singh, Kiran Adsul, “Flight Fare Prediction using Random Forest Algorithm”, International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Volume 2, Issue 3, ISSN (Online) 2581-9429, May 2022.
[13] Tadvi Shabana, Khan Mariya, Shaikh Afifa, Sayyed Naziya Begum, “A NOVEL DATA SCIENCE AND ML APPROACH TO PREDICT AIRFARE”, JETIR, Volume 6, Issue 4, ISSN-2349-5162, April 2019.
[14] Abhijit Boruah, Kamal Baruah, Biman Das, Manash Jyoti Das, Niranjan Borpatra Gohain, “A Bayesian Approach for Flight Fare Prediction Based on Kalman Filter”, Progress in Advanced Computing and Intelligent Engineering, Volume 2, ICACIE 2017.
[15] R. R. Subramanian, M. S. Murali, B. Deepak, P. Deepak, H. N. Reddy and R. R. Sudharsan, "Airline Fare Prediction Using Machine Learning Algorithms, " 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022, pp.877-884, doi: 10.1109/ICSSIT53264.2022.9716563.
[16] S. N. Prasath, S. Kumar M and S. Eliyas, "A Prediction of Flight Fare Using K-Nearest Neighbors, " 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp.1347-1351, doi: 10.1109/ICACITE53722.2022.9823876.
[17] Y. S. Can, K. Büyükoğuz, E. B. Giritli, M. Şişik and F. Alagöz, "Predicting Airfare Price Using Machine Learning Techniques: A Case Study for Turkish Touristic Cities, " 30th Signal Processing and Communications Applications Conference (SIU), doi: 10.1109/SIU55565.2022.9864692, 2022.
[18] G. Ratnakanth, "Prediction of Flight Fare using Deep Learning Techniques, " 2022 International Conference on Computing, Communication and Power Technology (IC3P), 2022, pp.308-313, doi: 10.1109/IC3P52835.2022.00071.
[19] Mrs. Sowjanya, Saud Ikram, Rayyan Khan, Shaik Haseeb Jawad, Saad ul arifeen, “Prediction of Airfare Prices Using Machine Learning”, International Journal of Mechanical Engineering, Vol.7 No.6, ISSN: 0974-5823, June, 2022.

[20] B. S. Panda, B. Phanendra Varma, B. Chandini, R. Bhoomika, "Flight Price Prediction Using Machine Learning Techniques", International Journal of Computer Sciences and Engineering, Vol.10, Issue.9, E-ISSN: 2347-2693, September 2022.
[21] Jaya Shukla, Aditi Srivastava, and Anjali Chauhan, "Airline Price Prediction using Machine Learning", International Journal of Research in Engineering, IT and Social Sciences, ISSN 2250-0588, Volume 10 Issue 05, May 2020.
[22] Vinal Raja, Janhavi Vakil, Yash Shah, Sonia Relan, "Prediction of Airfare Using Machine Learning", IJSDR, Volume 3, Issue 4, ISSN: 2455-2631, April 2018.
[23] Bhavuk Chawla, Ms. Chandandeep Kaur, "Airfare Analysis And Prediction Using Data Mining And Machine Learning", International Journal of Engineering Science Invention, Volume 6 Issue 11, ISSN (Online): 2319-6734, November 2017.
[24] Benny Mantin and Eran Rubin, "Fare Prediction Websites and Transaction Prices: Empirical Evidence from the Airline Industry", INFORMS, Vol.35, No.4, pp.640-655, July-August 2016.
[25] Mohit Vyavhare, Komal Wable, Ayush chothe and Shreya Kute, "Airfare Price Prediction Using Machine Learning Algorithms", IJARIIE, Vol-8 Issue-2,-ISSN (O)-2395-4396, 2022