



Flight Fare Prediction Using Data Science

¹Vaidehi Jadhav, ²Dr. Umesh Kulkarni

¹Master of Engineering Student, ²Master of Engineering Professor

¹Computer Engineering,

¹Vidyalankar Institute of Technology, Mumbai, India

Abstract : Air travel is a critical component of achieving economic growth and development. On a national, regional, and international scale, air travel fosters integration into the global economy and provides crucial linkages. It contributes to the growth of trade, tourism, and job possibilities for the same flight, even for close seats within the same cabin, airline ticket costs can now vary dynamically and significantly. Customers are looking for the best deal, while airlines are attempting to optimise their profit by maintaining a high overall revenue from a consumer's standpoint, determining the ideal time to procure plane tickets is difficult, owing to the lack of information necessary to make informed decisions regarding future price changes. The sole motive of this study is to use historical data to find underlying trends in airline pricing in India and to recommend the optimum time to purchase a flight ticket

IndexTerms - Data Science, Python, Prediction analysis, Jupyter Note book.

I. INTRODUCTION

This Demand prediction and pricing discrimination are two examples of computational approaches used by airlines to increase revenue. Different academics have proposed two types of models to help customers save money: models that forecast the optimal moment to acquire a ticket and models that anticipate the cheapest ticket price. The path, disembarkation month, disembarkation day, hour of disembarkation, if the day of disembarkation is a holiday, and aviation carrier all influence pricing trends. Highly competitive routes, such as most business trips, exhibited a non-decreasing pattern in which costs grew as the quantity of days till departure reduced, but other routes had a defined time frame during which prices were at their lowest. The flight fare dataset was obtained from Kaggle for this project. We must perform data preprocessing on the supplied data, which comprises airline, journey date, source, destination, route, departure time, arrival time, and duration. The departure and arrival times are also extremely important. Numpy, pandas, matplotlib, and seaborn are among the python libraries I've used.

II. RELATED WORK

Below are paper referred in sense of data science and prediction analysis of various datasets

A. Utilizing a Data analytics project to increase student awareness of contemporary global issues. [1]

This research demonstrated how a project in a Compositional Thinking class might assist students become more aware of current global concerns by providing and stimulating a variety of opportunities. The possibilities for expanding this project to other areas of interest and to a higher level are endless.

B. Data Science Framework - Heart Disease Predictions, Variant Models and Visualizations. [2]

The approach trial evidenced that the support vector machine and logistic regression generated the best classification results for the heart sickness dataset. The purpose is to look into how algorithms are used to categorize data in order to analyze and forecast cardiac disease.

C. Design and Implementation of Domestic News Collection System Based on Python. [3]

This system makes every attempt to make analyzing news information as simple as possible for users, and it offers news from various of websites. Because of its simple design, users can understand fully the news since it scrapes and provides the most meaningful data while disregarding the rest.

D. Prediction of House Pricing Using Machine Learning with Python. [4]

Multiple algorithms are utilized in this study to forecast the sale price of apartments. Task was completed using Python data mining techniques and algorithms to obtain the findings. The house sale prices were calculated meticulously. Data visualization was used to render a number of charts that depicted the distribution of data in various forms.

III. EXISTING SYSTEM

The intention could be to estimate house prices in Ames, Iowa at various locations, and the elements that influence house prices, as in the multitude of floors, garden space, total area, and so on, for a total of 79 variables. The datasets came from the Kaggle repositories. To forecast the charge of a home, the Python programming language and the method of Cognitive Computing were utilized. [4]

IV. PROPOSED SYSTEM

The target with this proposal is to improve the prediction model’s performance by examining the functionality of the features and removing any that aren’t needed. Finally, I use the features I’ve chosen to render prediction model, which produces an output value in the configuration of forecasted airticket price. Information regarding airline traffic and passenger volume for each market segment is needed to create the airticket price model at the market segment level. As a result, our suggested approach makes use of two accessible datasets.

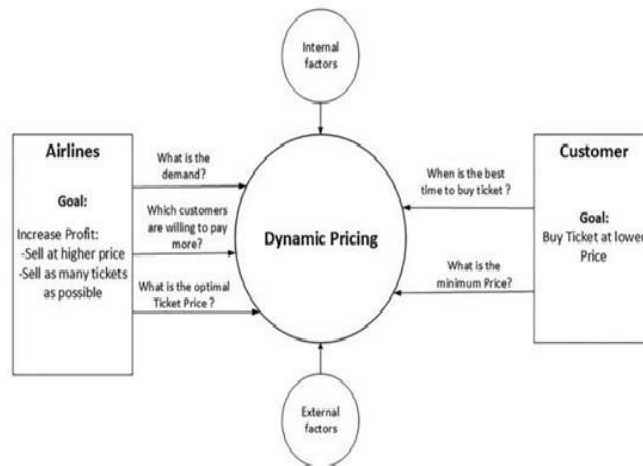


Figure.1 Proposed System

V. METHODOLOGY

This project’s methodology is implemented in a variety of ways, as detailed below.

A. Algorithm Design

A straightforward linear statistical analysis is used to find the causative link between two constant variables. The factor predictive is one of this same half variables. It provides a statistical relationship rather than a deterministic one. Decision tree analysis is engaged in which, the tree count system divides the acquired data into discrete sections while also rendering it permanent. The decision-making concentrates and leaf centers, are visible in the final findings. At the least, this decision-making hub may have two branches. Random forest technique is used combines the less accurate framework or model with the more predictor model to obtain improved inferential models. It combines the fundamental template into a larger model. To obtain decision trees that are highly nonstationary, the characteristics are tested and scattered to trees without replacement.

B. Data Design

Data collection is the systematic acquisition and quantification of information on variables of interest that allows researchers to answer research questions, test hypotheses, and evaluate outcomes. Data collecting is used in all sectors of study, including sciences, economics, business, and others. The data design flow shows the process in this the data is been used in order to the analysis.

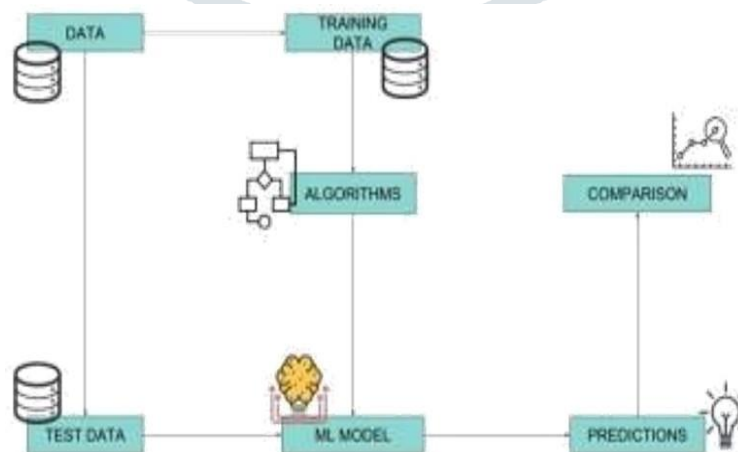


Figure.2 Data Design

VI. IMPLEMENTATION

Implementation process is carrying out, execution, or practice of a plan, technique for doing prediction of this tactic to retrieve the desired output.

Feature selection is a technique for limiting the input variable to your model using only useful data and eliminating noise. It’s the process of selecting appropriate characteristics for your machine learning model based on the sort of problem you’re attempting to answer automatically. Two data set are used in this project which performs the feature selection and give result to plot the data into histogram.

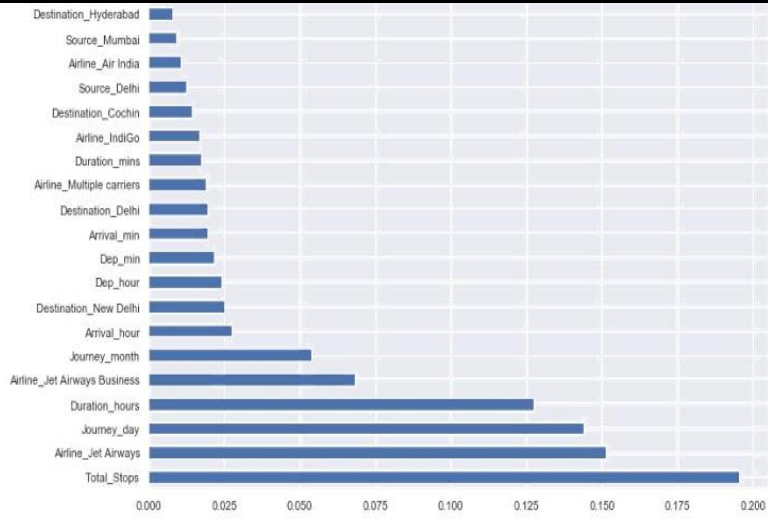


Figure.3 Feature Selection

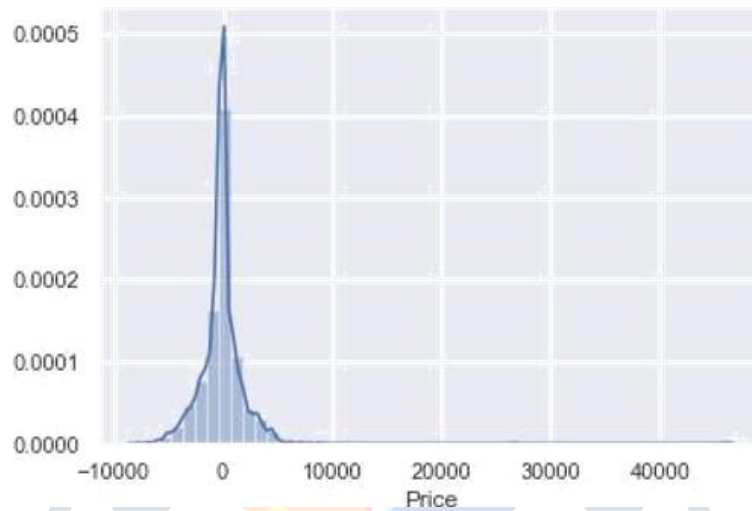


Figure.4 Histogram Plot

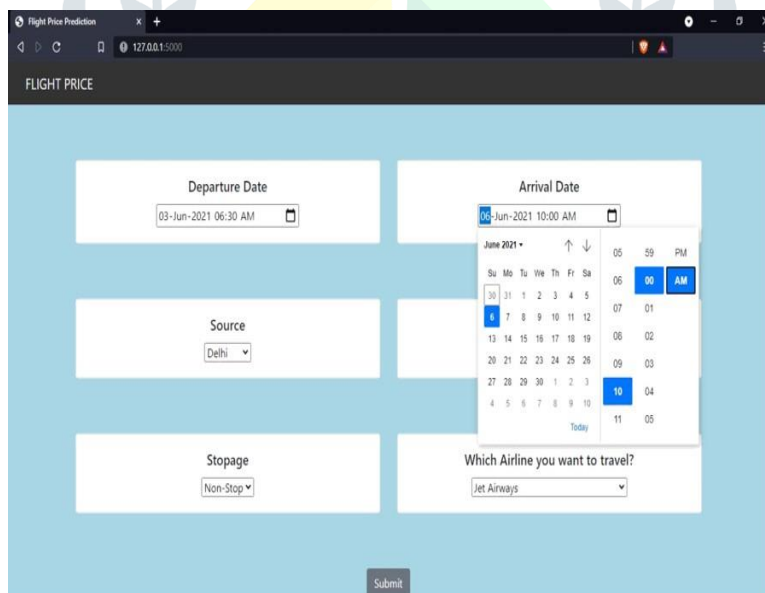


Figure.5 Selection Process (1)

Figure.6 Selection Process (2)

With the result displayed on the web application, the interfacing of datasets and applied features on data is finished. The web application allows the user to choose their preferred arrival and departure dates and times. Users are given complete control over the flight's origin and destination. The stop and nonstop function in the online application allows users to choose between a straight flight (which is achieved by the "non stop" feature) and a number of stop locations. The flight of the desired aviation firm can be selected from the drop down list that shows. The user can click the submit button to retrieve the results after making the suitable selection from the drop down list. As a result, our project assists the user in obtaining a flight fare in a matter of seconds. Using the features in the user interface, the user can acquire the desired output in real time and date.

Figure.7 Price Results

VII. CONCLUSION

Consumers find it challenging to determine the ideal time to buy aircraft tickets, mostly because they lack the knowledge necessary to make informed decisions about future price fluctuations. My major goal for this project was to leverage historical data to uncover underlying trends in airline pricing in India, as well as to suggest the best period to buy an airfare. The airline, the date of the journey, the source, the destination, the route, the departure time, the arrival time, and the duration are all included in the dataset. We must execute data pre-processing on the provided data. Some of the python libraries I've used are numpy, pandas, matplotlib, and seaborn. We were able to figure out which characteristics have the greatest impact on airfare premised on the prognostication.

VIII. ACKNOWLEDGMENT

This venture could not be possible without cooperation of the computer department. Alone without tremendous support and cooperation of a great amount of persons, it couldn't be possible. I'd also want to thank my project guide, Dr. Umesh Kulkarni Sir, for his suggestions.

REFERENCES

- [1] Rejitha Ravikumar, Pushpa Sadhwani, “Utilizing a Data analytics project to increase student awareness of contemporary global issues” Higher Colleges of Technology, Dubai, U.A.E., 2020.
- [2] Chittampalli Sai Prakash, Myneni Madhu Bala, Attluri Rudra, “DataScience Framework - Heart Disease Predictions, Variant Models and Visualizations” Institute of Aeronautical Engineering, Hyderabad, Telangana, India., 2020.
- [3] Haixia Lv, “Design and Implementation of Domestic News Collection System Based on Python” Shandong University of Finance and Economics Jinan China., 2020.
- [4] Namit Jain, Parikshay Goel, Purushottam Sharma, Vikas Deep “Prediction of House Pricing Using Machine Learning with Python” Department of Information Technology, Sector 125, Amity University Uttar Pradesh, Noida, India., 2020.
- [5] P. Malighetti, S. Paleari and R. Redondi, “Pricing strategies of low-cost airlines: The Ryanair case study,” Journal of Air Transport Management, vol. 15, no. 4, pp. 195-203, 2009.
- [6] 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.
- [7] W. Groves and M. Gini, “A regression model for predicting optimal purchase timing for airline tickets,” Technical Report 11-025, University of Minnesota, Minneapolis, 2011.
- [8] 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.
- [9] R. Ren, Y. Yang and S. Yuan, “Prediction of airline ticket price,” Technical Report, Stanford University, 2015.
- [10] M. Papadakis, “Predicting Airfare Prices,” 2014.T. Janssen, “A linear quantile mixed regression model for prediction of airline ticket prices,” Bachelor Thesis, Radboud University, 2014.

