

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

TEDIA

International Journal of Innovative Research in SCIENCE | ENGINEERING | TECHNOLOGY

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

## Impact Factor: 8.118

9940 572 462

6381 907 438

 $\bigcirc$ 



International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)



| e-ISSN: 2319-8753, p-ISSN: 2320-6710| <u>www.ijirset.com</u> | Impact Factor: 8.118|

Volume 11, Issue 6, June 2022

| DOI:10.15680/IJIRSET.2022.1106228 |

### **Flight Price Prediction using Machine Learning**

Namdev Gurav, Vinayak Lohar, Namdev Khandekar, Akash Rananaware

U.G. Student, Department of Computer Science and Engineering, JSPM's Bhagwant Institute Of Technology ,Barshi

Solapur, Maharashtra, India<sup>1</sup>

**ABSTRACT**: Optimal timing for airline ticket purchasing from the consumer's perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. In this project we majorly targeted to uncover underlying trends of flight prices in India using historical data and also to suggest the best time to buy a flight ticket. For this project, we have collected data from 18 routes across India while the data of 4 routes were extensively used for the analysis due to the sheer volume of data collected over 4 months resulting in 5.28 lakh data points each across the Mumbai-Delhi and Delhi-Mumbai route and 1.05 lakh data points each across the Delhi-Guwahati and Guwahati-Delhi route. The project implements the validations or contradictions towards myths regarding the airline industry.

**KEYWORDS**: Flight Price Prediction using Machine learning, **Flight Price Prediction for Easy Flight Booking**, **Book Your Flight, Price Predictor**.

#### I. INTRODUCTION

Since the deregulation of the airline industry, airfare pricing strategy has developed into a complex structure of sophisticated rules and mathematical models that drive the pricing strategies of airfare [1] [2] [3]. Although still largely held in secret, studies have found that these rules are widely known to be affected by a variety of factors [4] [5]. Traditional variables such as distance, although still playing a significant role, are no longer the sole factor that dictate the pricing strategy. Elements related to economic, marketing and societal trends have played increasing roles in dictating the airfare prices. Most studies on airfare price prediction have focused on either the national 2level or a specific market. Research at the market segment level, however, is still very limited. We define the term market segment as the market/airport pairbetween the flight origin and the destination. Being able to predict the airfare trend at the specific market segment level is crucial for airlines to adjust strategy and resources for a specific route. However, existing studies on market segment price prediction use heuristic-based conventional statistical models, airfare price. More specifically, our proposed framework extracts information from two specific public datasets, the DB1B and the T-100 datasets that are collected and maintained by the Office of AirlineInfor- mation within the United States Bureau of Transportation Statistics (BTS). The DB1B dataset has been utilized in various studies that assess the determinants of aircraft char- acteristics and frequency of flights [13], analyses for the structure and dynamics of O-D for the core of the air travel market [14], and demand-prediction [15]. The T-100 dataset includes air 3passenger volumes for U.S. domestic and international markets and covers large certified carriers that hold Certificates of Public Convenience and Necessity. The goal of our proposed framework is to draw a comprehensive profile of each market and uses machine learning techniquesto predict the average airfare on Market segment level.

#### **II. IDENTIFY, RESEARCH AND COLLECT IDEA**

In this project We are creating a Machine learning model to predict the price of flight .Anyone who has booked a flight ticket knows how unexpectedly the prices vary. Airlines use using sophisticated quasi-academic tactics known as "revenue management" or "yield management". The cheapest available ticket for a given date gets more or less expensive over time.

#### International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)



| e-ISSN: 2319-8753, p-ISSN: 2320-6710| <u>www.ijirset.com</u> | Impact Factor: 8.118|

Volume 11, Issue 6, June 2022

| DOI:10.15680/IJIRSET.2022.1106228 |

#### **III. STUDIES AND FINDINGS**

Data For any prediction/classification problem, we need historical data to work with. In this project, past flight prices for each route needs to be collected on a daily basis.Manually collecting data daily is not efficient and thus a python script was run on a remote server which collected prices daily at specific time.

#### **IV. EXPERIMENTAL RESULTS**

Figures shows the results of text detection from an image and inpainting by using exemplar based Inpainting algorithm. Figs. 2, 3, 4 (a) shows the original image. (b) is the image obtained by applying first set of criteria. In detailed analysis for the Delhi - Guwahati Route

The trends in the data collected for the sector of Delhi to Guwahati busted some of the very famous myths assumed by travellers of the aviation industry.1. Flight prices do not increase continuously as the Date of Departure approaches closer 2 : Flight Price vs Days to Departure With the validation of the problem statement and with a scope to predict when to buy and when to wait, we begin the analysis of the dataset. The dataset of the flight prices follows a Lognormal distribution with ome outliers which have been ignored as we are only interested with the minimum fare corresponding to a certain route.

df = pd.read_excel(r'D:	\project_python\Data_T	<pre>[rain_Flight.xlsx')</pre>
df.shape		

(10683, 11)

df.head()											
	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
0	IndiGo	24/03/2019	Banglore	New Delhi	$BLR \rightarrow DEL$	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
1	Air India	1/05/2019	Kolkata	Banglore	$CCU \to IXR \to BBI \to BLR$	05:50	13:15	7h 25m	2 stops	No info	7662
2	Jet Airways	9/06/2019	Delhi	Cochin	$DEL \to LKO \to BOM \to COK$	09:25	04:25 10 Jun	19h	2 stops	No info	13882
3	IndiGo	12/05/2019	Kolkata	Banglore	$CCU \to NAG \to BLR$	18:05	23:30	5h 25m	1 stop	No info	6218
4	IndiGo	01/03/2019	Banglore	New Delhi	$BLR \rightarrow NAG \rightarrow DEL$	16:50	21:35	4h 45m	1 stop	No info	13302

plt.figure(figsize = (15,8))
sns.distplot(data.Price)

<matplotlib.axes.\_subplots.AxesSubplot at 0x23fd9528bb0>



#### International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)



| e-ISSN: 2319-8753, p-ISSN: 2320-6710| <u>www.ijirset.com</u> | Impact Factor: 8.118|

Volume 11, Issue 6, June 2022

| DOI:10.15680/IJIRSET.2022.1106228 |

#### V. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions

#### REFERENCES

- 1. O. Etzioni, R. Tuchinda, C. A. Knoblock, and A. Yates. To buy or not to buy: mining airfare data to minimize ticket purchase price.
- 2. Manolis Papadakis. Predicting Airfare Prices.
- 3. Groves and Gini, 2011. A Regression Model For Predicting Optimal Purchase Timing For Airline Tickets.
- 4. Modeling of United States Airline Fares Using the Official Airline Guide (OAG) and Airline Origin and Destination Survey (DB1B), Krishna Rama-Murthy, 2006.
- B. S. Everitt: The Cambridge Dictionary of Statistics, Cambridge University Press, Cambridge (3<sup>rd</sup> edition, 2006). ISBN 0-521-69027-7.
- 6. Bishop: Pattern Recognition and Machine Learning, Springer, ISBN 0-387-31073-8.
- 7. E. Bachis and C. A. Piga. Low-cost airlines and online price dispersion. International Journal of Industrial Organization, In Press, Corrected Proof, 2011.
- 8. Y. Levin, J. McGill, and M. Nediak. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. Management Science, 55(1):32–46, 2009.





Impact Factor: 8.118





## INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com