A Project Report on

OPTIMIZING FLIGHT BOOKING DECISION :

Through Machine learning price prediction.

by

Team Leader: PULIMI SUSHMA (20AT1A05G8)
Student 1 : MALA SWETHESWARI DEVI(20AT1A05H1)

Under the Guidance of

MRS.M. JAYA SUNITHA, MTech

Associate Professor

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

G. PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY (Autonomous)

(Approved by AICTE | NAAC Accreditation with 'A' Grade | Accredited by NBA (ECE, CSE, EEE, CE) | Permanently Affiliated to JNTUA)

ABSTRACT

The Flight ticket prices increase or decrease every now and then depending on various factors like timing of the flights, destination, duration of flights. In the proposed system a predictive model will be created by applying machine learning algorithms to the collected historical data of flights. 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. The project implements the validations or contradictions towards myths regarding the airline industry, a comparison study among various models in predicting the optimal time to buy the flight ticket and the amount that can be saved if done so. Remarkably, the trends of the prices are highly sensitive to the route, month of departure, day of departure, time of departure, whether the day of departure is a holiday and airline carrier. Highly competitive routes like most business routes (tier 1 to tier 1 cities like Bangalore-Delhi) had a non-decreasing trend where prices increased as days to departure decreased, however other routes (tier 1 to tier 2 cities like Bangalore - calcutta) had a specific time frame where the prices are minimum. Moreover, the data also uncovered two basic categories of airline carriers operating in India – the economical group and the luxurious group, and in most cases, the minimum priced flight was a member of the economical group. The data also validated the fact that, there are certain time-periods of the day where the prices are expected to be maximum. The scope of the project can be extensively extended across the various routes to make significant savings on the purchase of flight prices across the Indian Domestic Airline market.

Contents

LIST OF FIGURES AND TABLES

Database collected from KAGGLE about the flight booking

And data consist of different features like date of journey,arrival time,departure time etc

CHAPTER 1

INTRODUCTION

The flight ticket buying system is to purchase a ticket many days prior to flight take-off so as to stay away from the effect of the most extreme charge. Mostly, aviation routes don't agree this procedure. Plane organizations may diminish the cost at the time, they need to build the market and at the time when the tickets are less accessible. They may maximize the costs. So, the cost may rely upon different factors. To foresee the costs this venture uses AI to exhibit the ways of flight tickets after some time. All organizations have the privilege and opportunity to change its ticket costs at any time. Explorer can set aside cash by booking a ticket at the least costs. People who had travelled by flight frequently are aware of price fluctuations. The airlines use complex policies of Revenue Management for execution of distinctive evaluating systems. The evaluating system as a result changes the charge depending on time, season, and festive days to change the header or footer on successive pages. The ultimate aim of the airways is to earn profit whereas the customer searches for the minimum rate. Customers usually try

to buy the ticket well in advance of departure date so as to avoid hike in airfare as date comes closer. But actually, this is not the fact. The customer may wind up by giving more than they ought to for the same seat.

Cleaning and Preparing of Data

Cleaning and preparing data are a very important step in machine learning**. The data collected can't**

RMSE

RMSE is a tool that helps in determining how accurately the model is making the predictions. It

calculates how much error the model creates while making these predictions. It measures the standard of

predictions. Mathematically, it is defined as the square root of the average of the squares of all the errors.

Error is defined as the difference between the actual and predicted value. Less the RMSE, the better the

performance of the model is. Usually, an RMSE score of less than 1 is considered the best.

LITERATURE REVIEW

Proposed study [1] Airfare price prediction using machine learning techniques, For the research work they have used dateset consisting of 1814 data flights of the Aegean Airlines collected and used to train machine learning model. Different number of features were used to train model various to showcase how selection of features can change accuracy of model. They have used various algorithms such as logistic regression, Generalized Regression, LLE(, Random Forest Regression Tree. Regression Tree, Regression SVM (Polynomial and Linear) and Linear Regression (LR) and gained different outputs for each machine learning algorithms. They have tried and trained various types of models with removing and

adding different features from the dateset. Followed typical data science life cycle. The best results came from Bagging regression tree.

2. William Groves and Maria Gini "An agent for optimizing airline ticket purchasing" in proceedings of the 2013 international conference on autonomous agents and multi-agent systems.

•

CHAPTER 3

PROPOSED METHOD

- As painful as it can seem, people actually enjoy planning their trips and can spend 2 / 4 hours glued to the screen to find the best place, the best itinerary, and the best price.
- And here's when Machine Learning and Artificial Intelligence come into play: by analyzing large datasets, the AI-infused travel systems can generate super personalized suggestions for the travellers.

CHAPTER 4

EXPERIMENTAL RESULTS

New Tab | flight-price-prediction.ipynb

https://colab.research.google.com/drive/1-y8biMfrZCemkcPaA7wmK2kQfAghdmTC

flight-price-prediction.ipynb

File Edit View Insert Runtime Tools Help Last saved at August 4

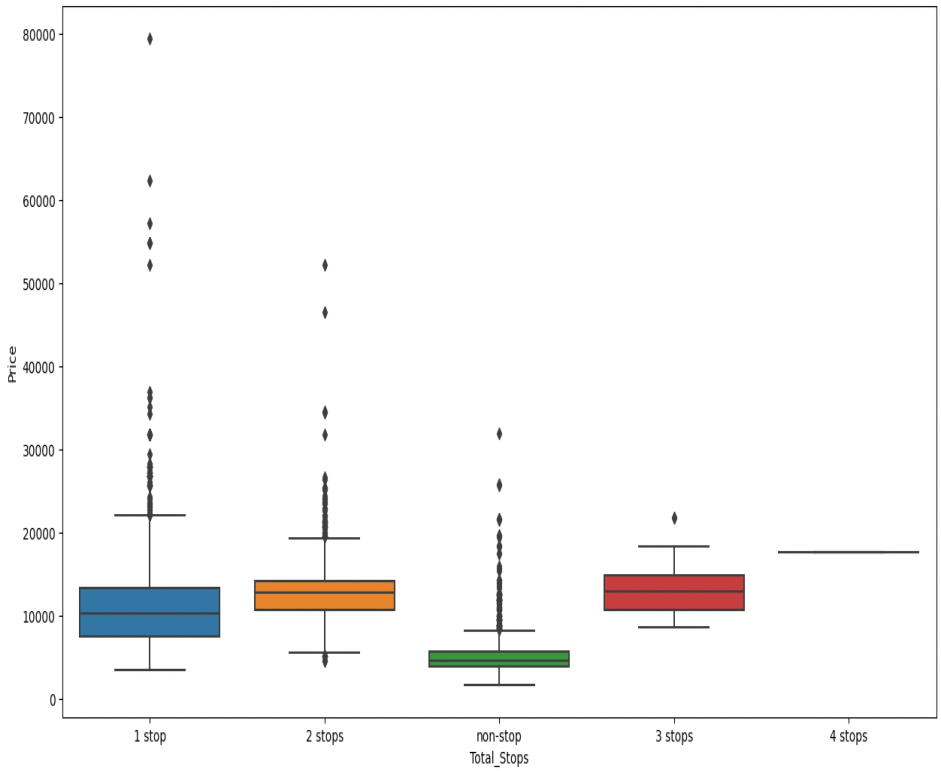
Comment Share Connect

+ Code + Text

Perform Total_Stops vs Price Analysis

```
[ ] plt.figure(figsize=(15,8))
    sns.boxplot(x='Total_Stops',y='Price',data=df.sort_values('Price',ascending=False))
```

```
<Axes: xlabel='Total_Stops', ylabel='Price'>
```



26°C

ENG

22:36
05-08-2023

New Tab | flight-price-prediction.ipynb

https://colab.research.google.com/drive/1-y8biMfrZCemkcPaA7wmK2kQfAghdmTC#scrollTo=impressed-istanbul

flight-price-prediction.ipynb

File Edit View Insert Runtime Tools Help Last saved at August 4

Comment Share

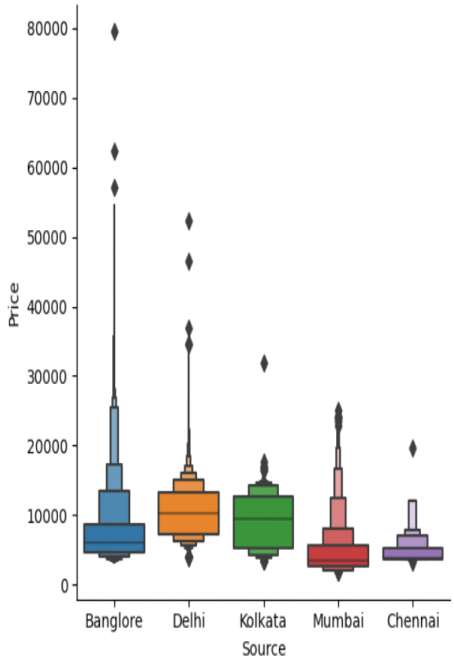
+ Code + Text

Connect

```
#Source vs Price

plt.figure(figsize=(15,15))
sns.catplot(x='Source',y='Price',data=df.sort_values('Price',ascending=False),kind='boxen')
```

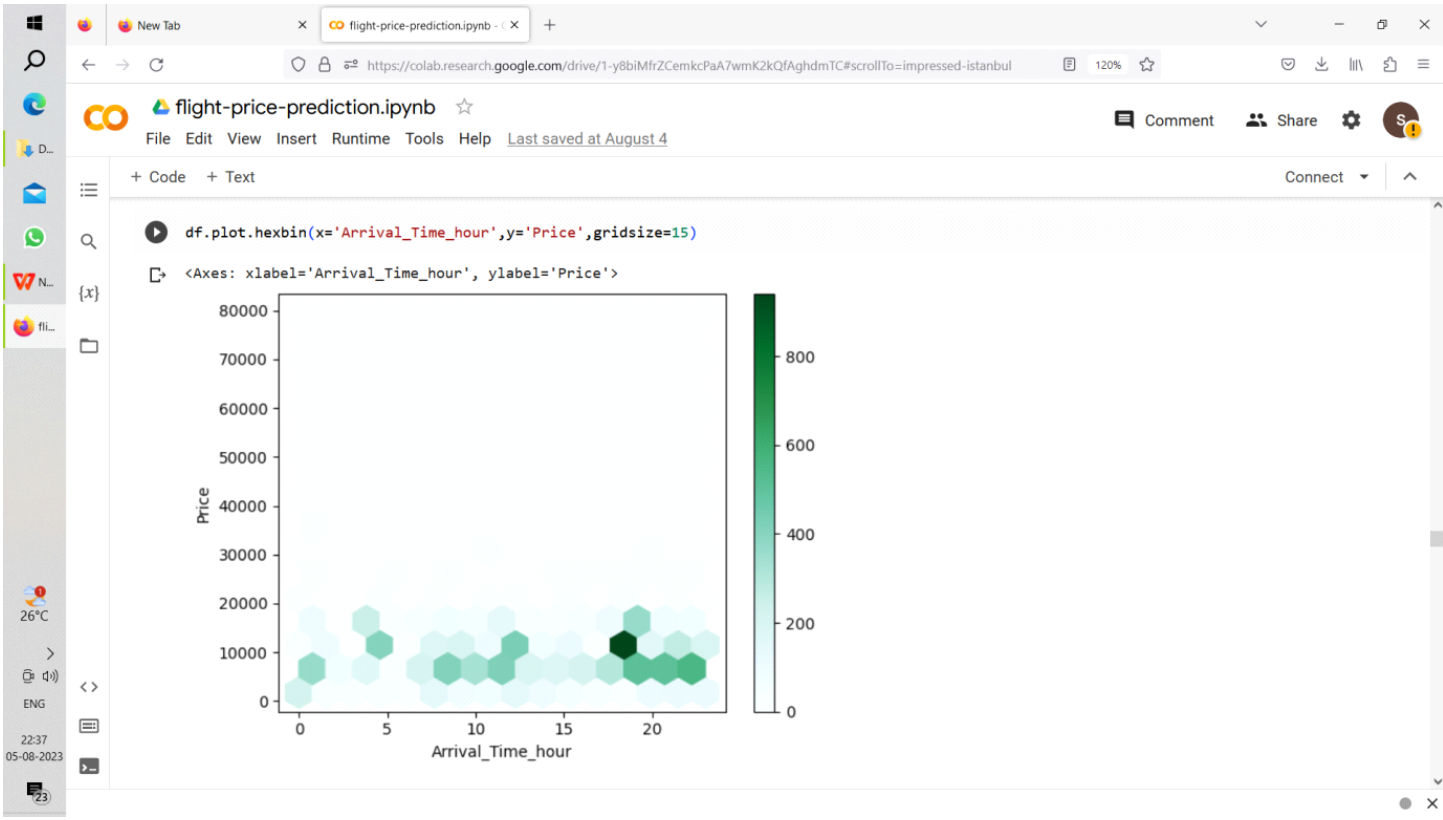
```
<seaborn.axisgrid.FacetGrid at 0x7def2efb6860>
<Figure size 1500x1500 with 0 Axes>
```

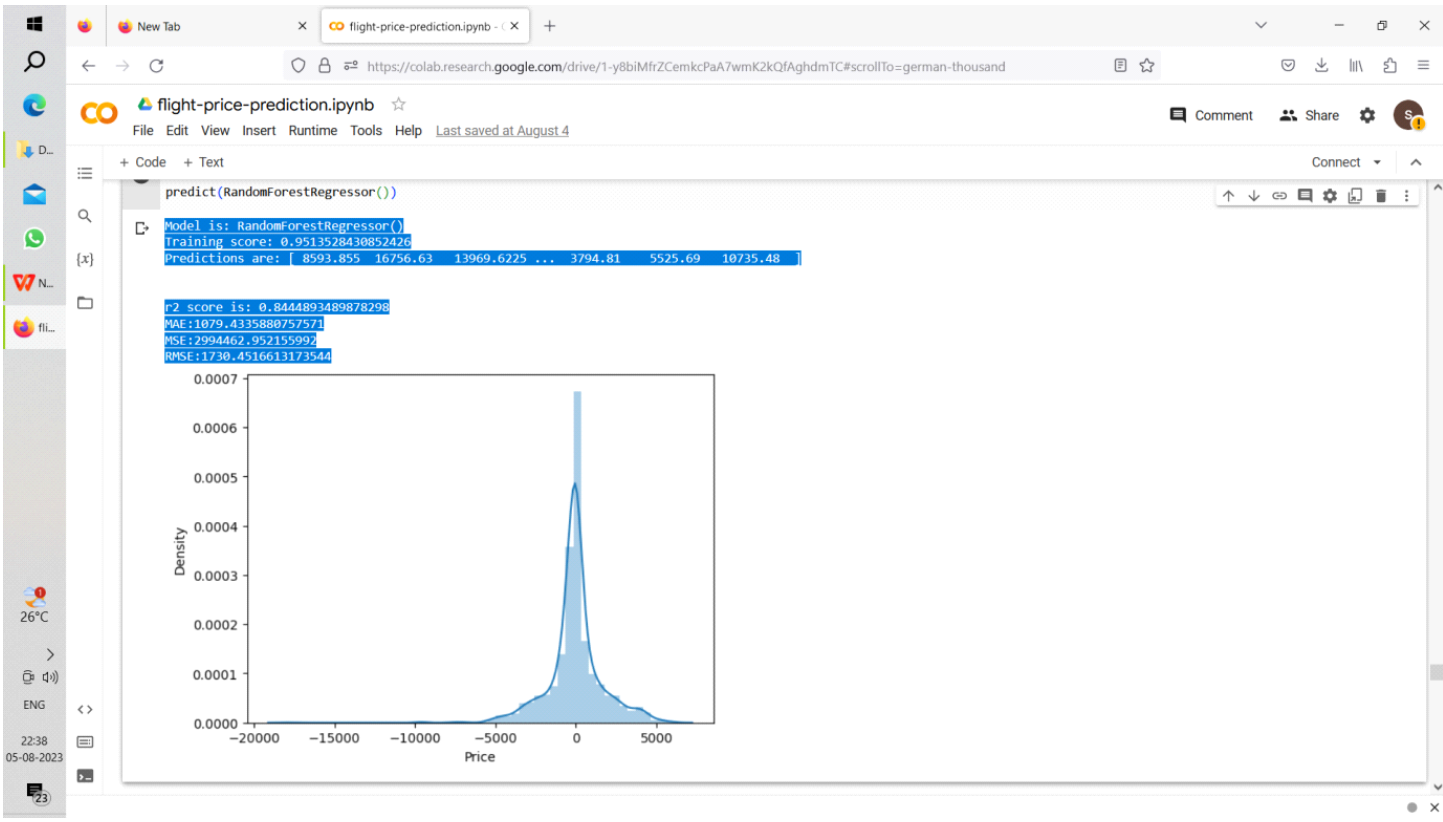


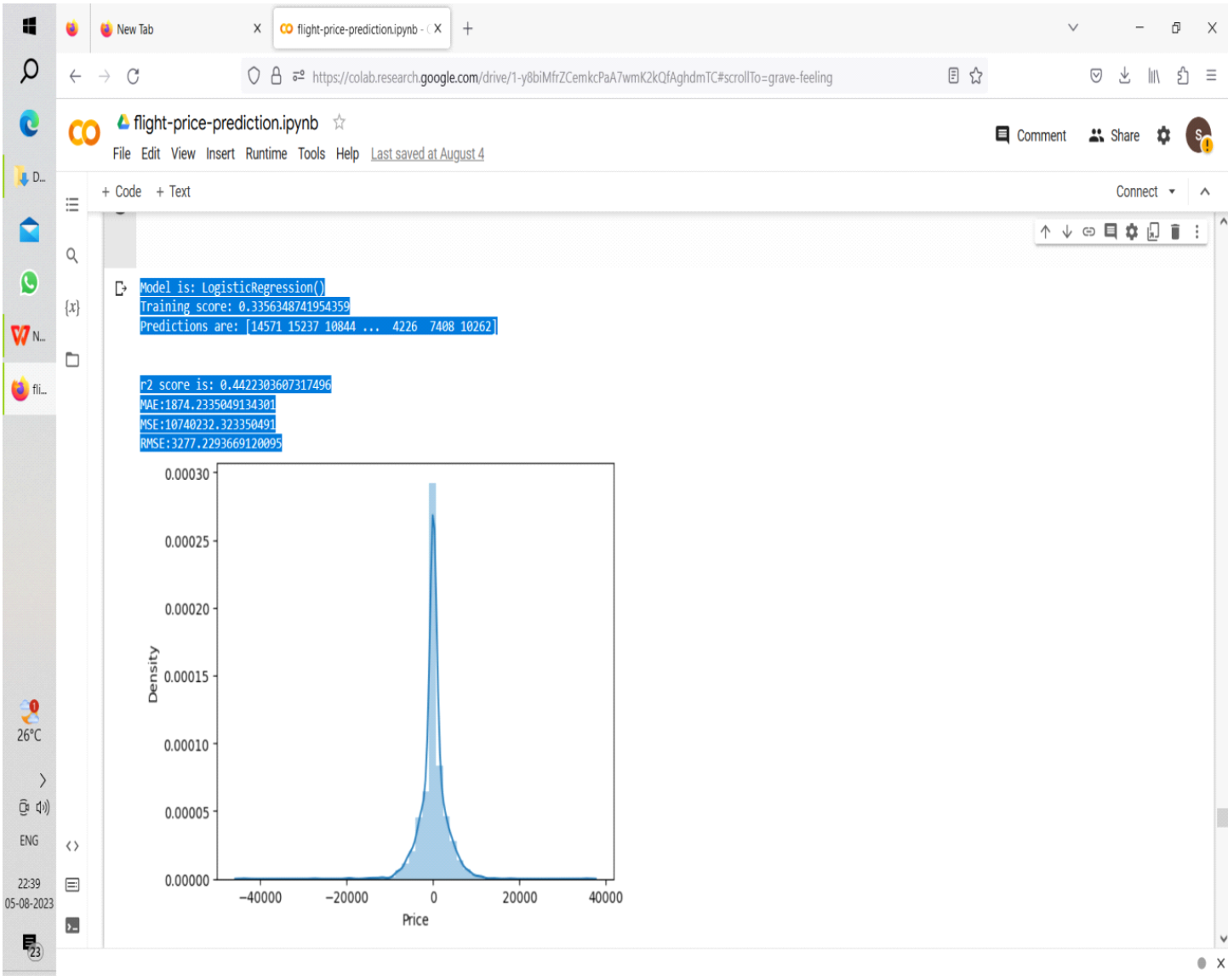
26°C

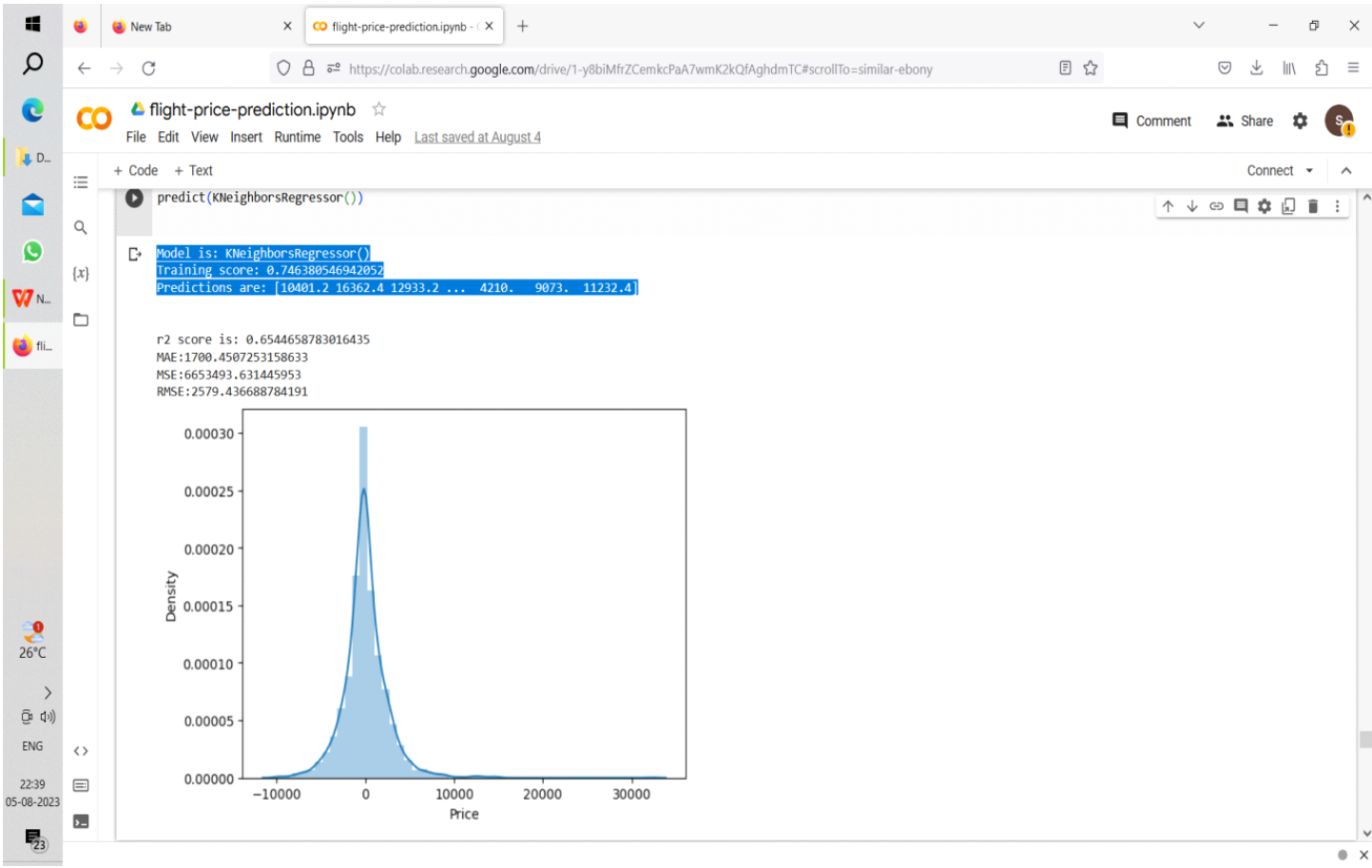
ENG

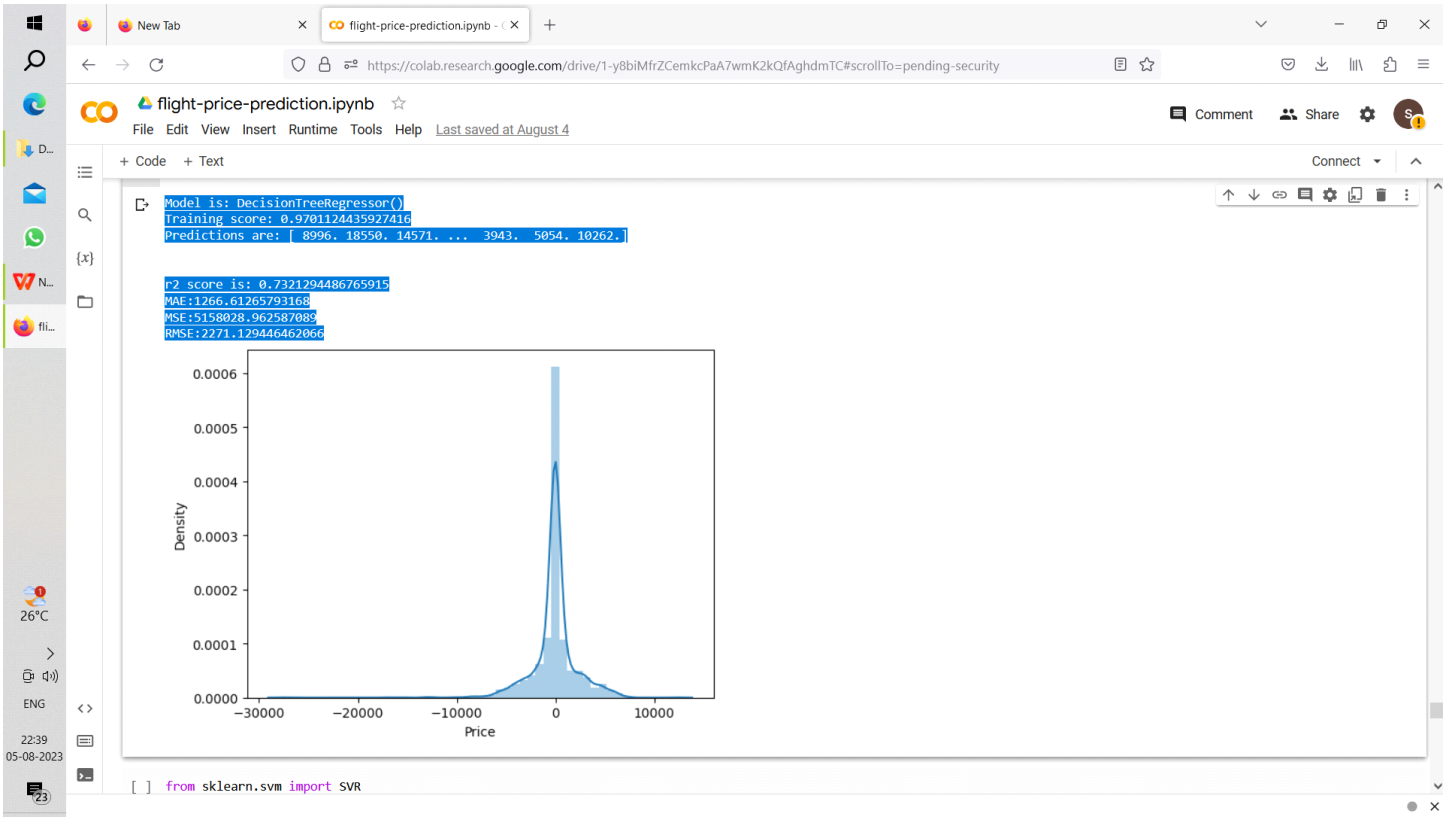
22:36
05-08-2023

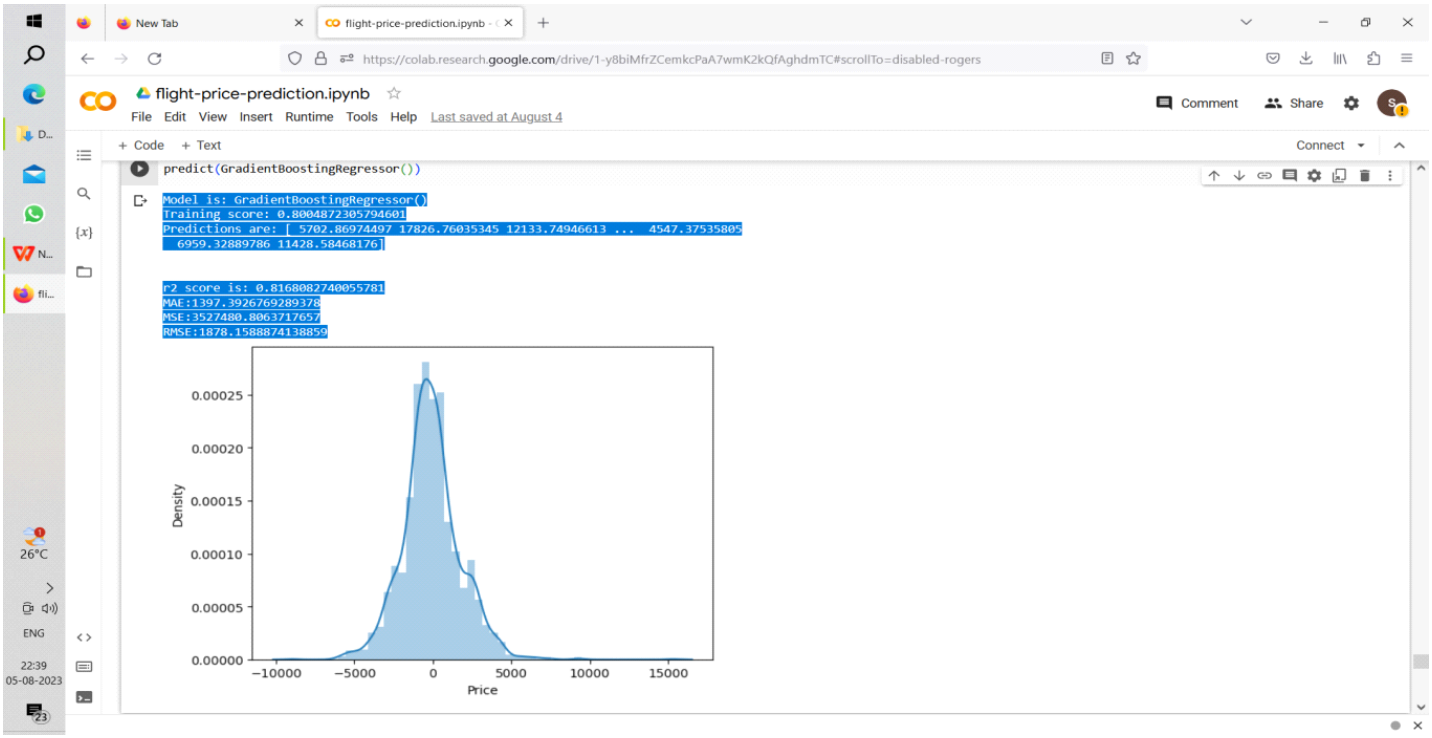

New Tab
flight-price-prediction.ipynb
https://colab.research.google.com/drive/1-y8biMfrZCemkcPaA7wmK2kQfAghdmTC#scrollTo=impressed-istanbul
flight-price-prediction.ipynb
File Edit View Insert Runtime Tools Help Last saved at August 4
Comment
Share
+ Code + Text
Connect
df.plot.hexbin(x='Arrival_Time_hour',y='Price',gridsize=15)
<Axes: xlabel='Arrival_Time_hour', ylabel='Price'>
Price
Arrival_Time_hour
80000
70000
60000
50000
40000
30000
20000
10000
0
0
5
10
15
20
800
600
400
200
0
26°C
ENG
22:37
05-08-2023

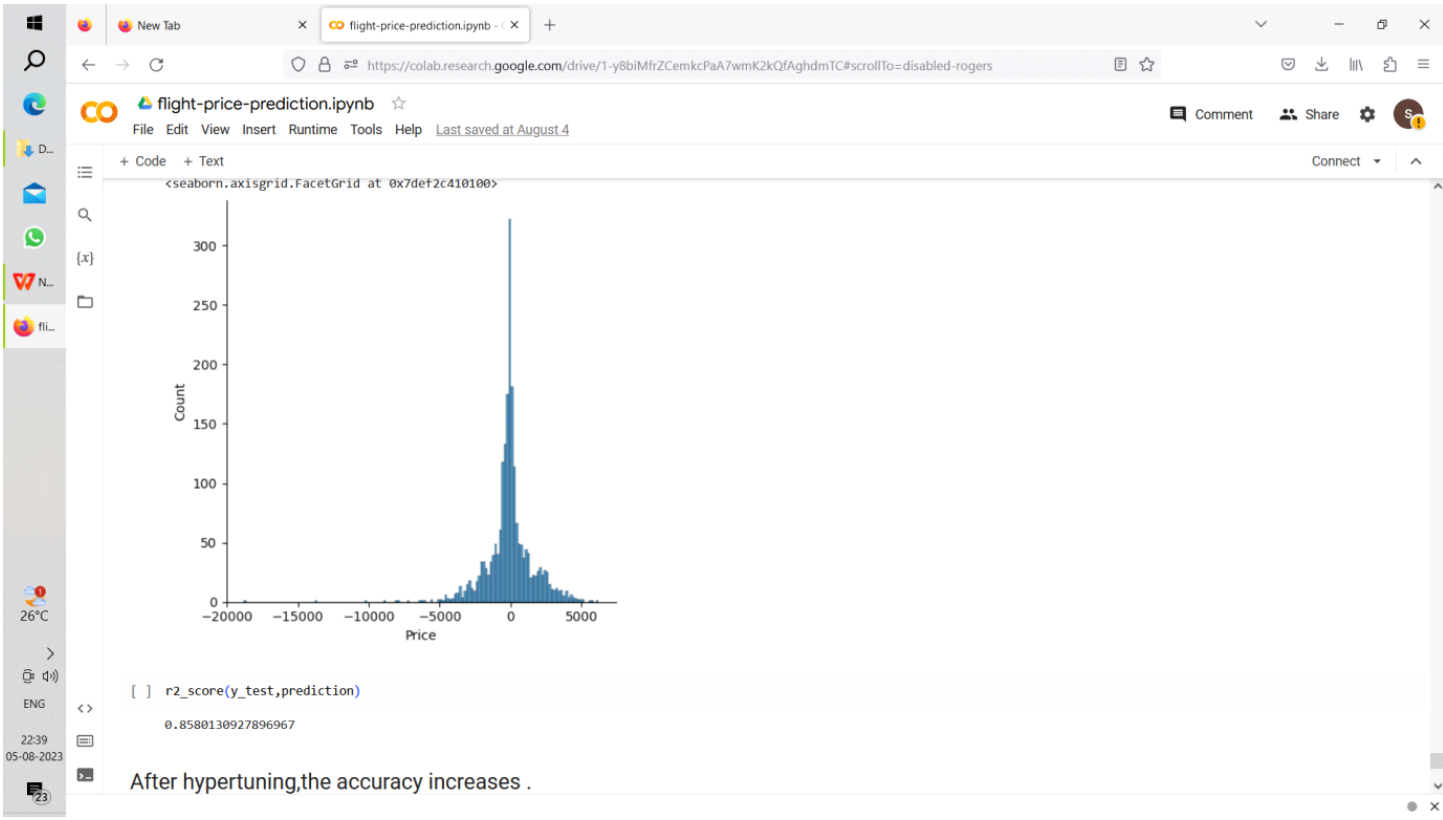

New Tab
flight-price-prediction.ipynb
https://colab.research.google.com/drive/1-y8biMfrZCemkcPaA7wmK2kQfAghdmTC#scrollTo=german-thousand
flight-price-prediction.ipynb
File Edit View Insert Runtime Tools Help Last saved at August 4
Comment
Share
+ Code + Text
Connect
predict(RandomForestRegressor())
Model is: RandomForestRegressor()
Training score: 0.9513528430852426
Predictions are: [8593.855 16756.63 13969.6225 ... 3794.81 5525.69 10735.48]
r2 score is: 0.8444893489878298
MAE:1079.4335880757571
MSE:2994462.952155992
RMSE:1730.4516613173544
Density
Price
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
−20000
−15000
−10000
−5000
0
5000
26°C
ENG
22:38
05-08-2023


New Tab
flight-price-prediction.ipynb
https://colab.research.google.com/drive/1-y8biMfrZCemkcPaA7wmK2kQfAghdmTC#scrollTo=grave-feeling
flight-price-prediction.ipynb
File Edit View Insert Runtime Tools Help Last saved at August 4
Comment
Share
+ Code + Text
Connect
Model is: LogisticRegression()
Training score: 0.3356348741954359
Predictions are: [14571 15237 10844 ... 4226 7408 10262]
r2 score is: 0.4422303607317496
MAE:1874.2335049134301
MSE:10740232.323350491
RMSE:3277.2293669120095
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
Density
−40000
−20000
0
20000
40000
Price
26°C
ENG
22:39
05-08-2023


New Tab
flight-price-prediction.ipynb
https://colab.research.google.com/drive/1-y8biMfrZCemkcPaA7wmK2kQfAghdmTC#scrollTo=similar-ebony
flight-price-prediction.ipynb
File Edit View Insert Runtime Tools Help Last saved at August 4
Comment
Share
+ Code + Text
Connect
predict(KNeighborsRegressor())
Model is: KNeighborsRegressor()
Training score: 0.746380546942052
Predictions are: [10401.2 16362.4 12933.2 ... 4210. 9073. 11232.4]
r2 score is: 0.6544658783016435
MAE:1700.4507253158633
MSE:6653493.631445953
RMSE:2579.436688784191
Density
Price
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
−10000
0
10000
20000
30000
26°C
ENG
22:39
05-08-2023


New Tab
flight-price-prediction.ipynb
https://colab.research.google.com/drive/1-y8biMfrZCemkcPaA7wmK2kQfAghdmTC#scrollTo=pending-security
flight-price-prediction.ipynb
File Edit View Insert Runtime Tools Help Last saved at August 4
Comment
Share
+ Code + Text
Connect
Model is: DecisionTreeRegressor()
Training score: 0.9701124435927416
Predictions are: [8996. 18550. 14571. ... 3943. 5054. 10262.]
r2 score is: 0.7321294486765915
MAE:1266.6126579316 8
MSE:5158028.962587089
RMSE:2271.129446462066
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
Density
−30000
−20000
−10000
0
10000
Price
[ ] from sklearn.svm import SVR
26°C
ENG
22:39
05-08-2023


New Tab
flight-price-prediction.ipynb
https://colab.research.google.com/drive/1-y8biMfrZCemkcPaA7wmK2kQfAghdmTC#scrollTo=disabled-rogers
flight-price-prediction.ipynb
File Edit View Insert Runtime Tools Help Last saved at August 4
Comment
Share
+ Code + Text
Connect
predict(GradientBoostingRegressor())
Model is: GradientBoostingRegressor()
Training score: 0.8004872305794601
Predictions are: [5702.86974497 17826.76035345 12133.74946613 ... 4547.37535805
6959.32889786 11428.58468176]
r2 score is: 0.816808274005781
MAE:1397.3926769289378
MSE:3527480.8063717657
RMSE:1878.1588874138859
Density
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
−10000
−5000
0
5000
10000
15000
Price
26°C
ENG
22:39
05-08-2023



CHAPTER 5

APPLICATIONS/ADVANTAGES

ADVANTAGES

One of the most significant benefits of AI in aviation is improved safety. AI can analyze vast amounts of data and identify potential safety issues before they become critical. For example, AI can/could monitor aircraft systems in real-time and detect anomalies that could indicate a problem. This could help prevent accidents and reduce maintenance costs.
AI can also improve operational efficiency in the aviation industry. By optimizing flight routes and schedules, airlines can reduce fuel consumption and save money. AI can also help airlines manage their fleets more effectively, by predicting maintenance needs and optimizing aircraft utilization.

Another area where AI can make a big impact is passenger experience. AI-powered chatbots and virtual assistants can provide passengers with personalized support throughout their journey, from booking to arrival. AI can also help airlines

predict passenger demand and adjust their offerings accordingly, enhancing customer satisfaction.

DIS ADVANTAGES

While AI has many potential benefits in aviation, it's important to be aware of the risks as well. One of the biggest concerns is the possibility of AI malfunctioning or being hacked. If an AI system fails, it could potentially cause a serious accident. Likewise, if an AI system is hacked, it could be used to gain unauthorized access to critical systems.

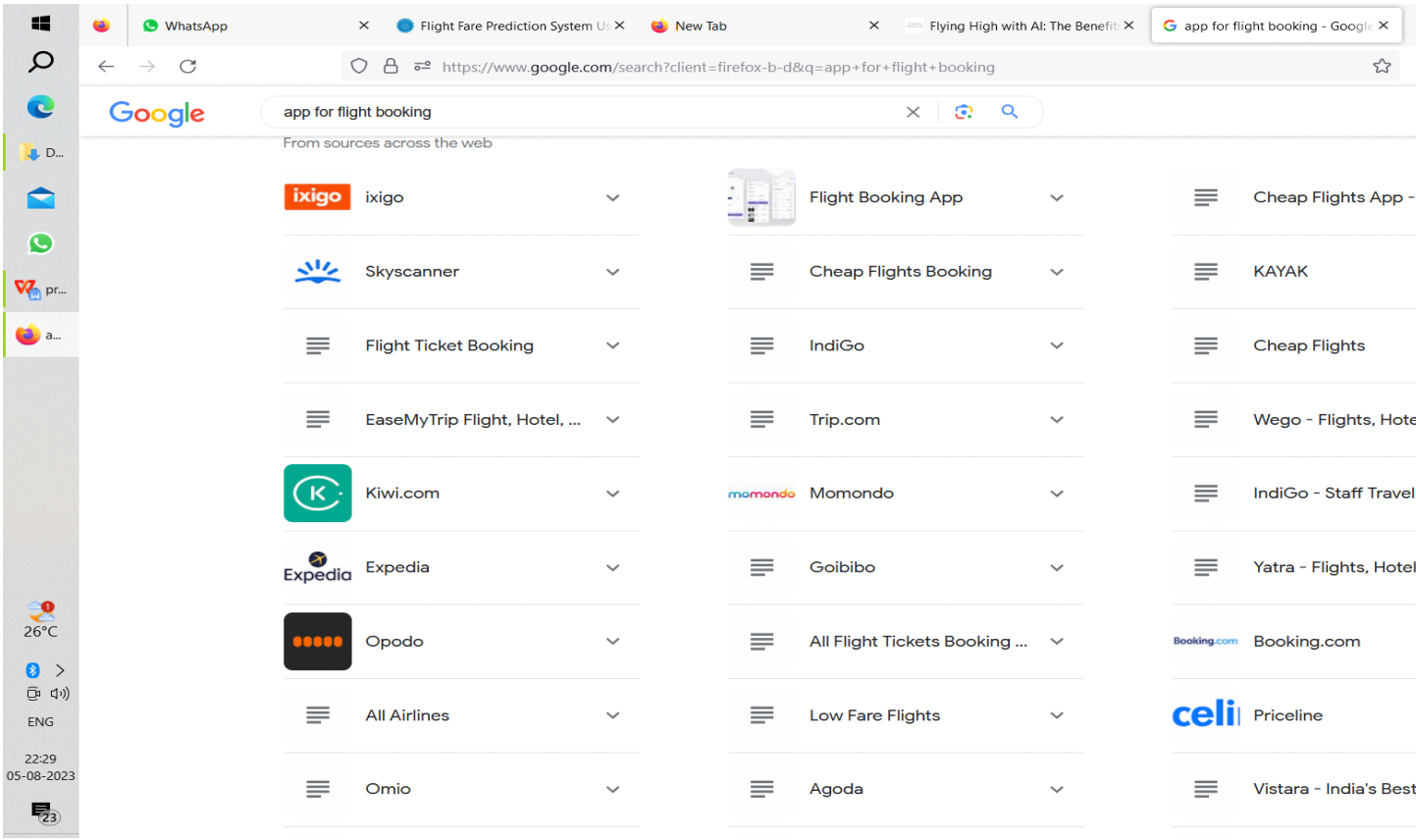
Another risk of AI in aviation is the potential for job loss. As AI becomes more prevalent in the industry, some jobs may become automated, putting human workers out of work. It's important to ensure that any automation is done responsibly and that human workers are not left behind.

APPLICATIONS

Every web-based-application that supports information publishing and sharing (text, video, audio, photo), the building of personal profiles, connecting to a world.

Applications / flight booking

From sources across the web



CHAPTER 6

CONCLUSIONS & FUTURE SCOPE

CONCLUSION

Machine Learning algorithms are applied on the dataset to predict the dynamic fare of flights. This gives the predicted values of flight fare to get a flight ticket at minimum cost. Data is collected from the websites which sell the flight tickets so only limited information can be accessed. The values of R-squared obtained from the algorithm give the accuracy of the model. In the future, if more data could be accessed such as the current availability of seats, the predicted results will be more accurate. Finally, we have created the entire process of predicting an airline ticket and given a proof of our predictions based on the previous trends with our prediction.

FUTURE SCOPE

AI is disrupting the airline industry in a number of ways, from **customer service and operations optimization to safety and security**. By leveraging the power of machine learning algorithms and other AI technologies, airlines can improve efficiency, reduce costs, and enhance the travel experience for their customers

-