Revolutionizing Air Travel through Predictive Intelligence

Spencer Dsheel

26/02/2024

Abstract

Being a rapidly evolving landscape, air travel plays a pivotal role in global connectivity and economic development. However, the volatility in flight prices poses a huge challenge to travellers. To overcome this dilemma, Artificial intelligence and Machine learning can be used to predict these fluctuations and forecast their trend to optimize the travel experience and ensure customer satisfaction. Focused on the unpredictable nature of flight prices, we sought to create a web app that not only forecasts prices but also encourages informed decision making capabilities. The web app integrates machine learning algorithm and providing users with accurate prediction and transparency.

1.0 Introduction

The commercial aviation industry is a crucial sector in the realm of modern transportation. It plays a significant role in enhancing connectivity and reducing the time taken to travel between destinations. Customers usually book their tickets months in advance to ensure that they get the cheapest prices for their travel as last-minute bookings are often expensive. Several aspects decide a flight price and customers find it confusing to fully understand these and get the lowest prices. People who travel frequently, especially those who enjoy backpacking and those who tend to book their flights on short notice, are always on the lookout for cheaper flights. They want to get the most value out of their travels while not breaking the bank.

Technologies such as Machine Learning and Deep Learning can be used to overcome this problem. Our report focuses on our Travel web app, which incorporates a Machine Learning algorithm to anticipate flight prices. The web app will also recommend whether customers should book their tickets or wait for a cheaper and more cost-effective price. Beyond being a technological advancement, this project represents a revolution where predictive intelligence meets the needs of modern travellers promising a more informed and personal journey

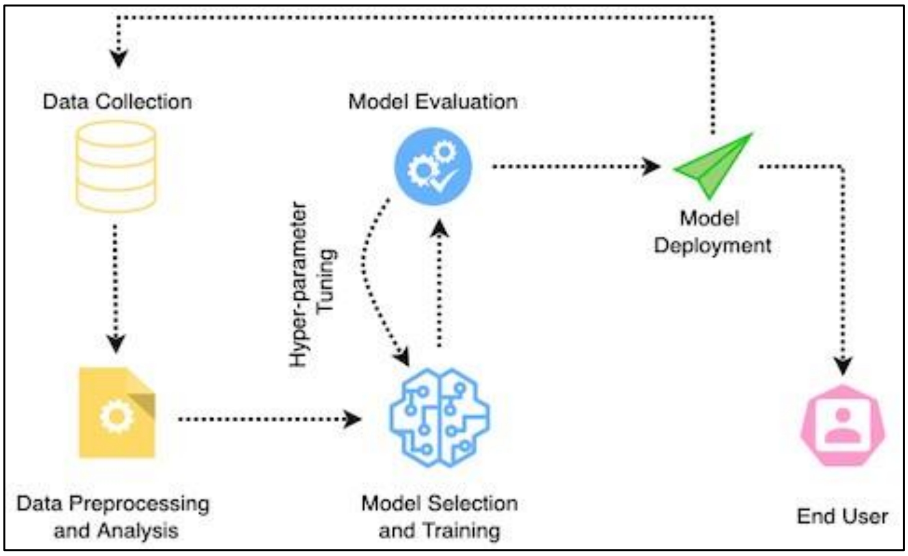

Figure 1: Model Diagram

2.0 Problem Statement

Presenting a technological approach to tackle the unpredictable fluctuations in flight prices by designing a travel web application that will accurately predict flight prices and make informed decisions for customers by lowering travel expenses.

3.0 Business Need Assessment

The aviation industry is always growing. IATA predicts that the number of passengers using flights will reach 4.7 billion in 2024. After the pandemic, the number of people travelling has increased and this responds well to the airlines. Passengers often search for affordable flight tickets during their travels, but predicting ticket prices can be challenging. The existing methods lack the sophistication needed to navigate the pricing fluctuations. This leads to a number of challenges for travellers ranging from financial strain and stress caused by the overwhelming number of options.

The aviation industry generates a massive amount of data every day. Machine learning models can utilize this data to predict the trend of flight prices. This, in turn, can be beneficial for passengers who make last-minute bookings. Our proposed web application aims to leverage this data and machine learning models to forecast flight prices and provide recommendations to passengers. The proposed model will utilize data from travel companies. Our web application will collaborate with travel agencies that book tickets daily which will help customers plan their trip accordingly. Our goal is to empower travellers with a tool that simplifies decision making and provides a reliable and user centric approach.

4.0 Target Specification and Characterization

Real time Updates

Incorporating real time data feeds will keep users updated about the latest changes in flight prices and availability.

Leveraging Artificial Intelligence and Machine Learning

Making use of Machine learning algorithms to buy affordable tickets and have an upper edge

Predicting flight prices and their trends using Machine Learning algorithms and historical data from airlines. If the algorithm predicts a downward trend while a customer is booking, the system will suggest that the customer wait. Based on the trend, the system will provide an approximate date when the ticket will be cheapest.

Cost Sensitivity

Travel agencies can make use of the prediction algorithm to book tickets for customers and plan their trip itinerary based on the recommendation of the web application. With these valuable insights, the user can plan trips that align with their budget.

5.0 External Search

This section will cover similar related work to our project.

Airline companies frequently adjust their ticket prices based on various factors, utilizing proprietary rules and algorithms in search of the most suitable pricing policy. In a recent report, [1] T. Kalampokas highlighted the use of Artificial Intelligence (AI) models for this task, given their compactness, fast adaptability, and potential for data generalization. This paper presents an analysis of airfare price prediction, with a focus on identifying similarities in the pricing policies of different airline companies using AI techniques [2].

Hopper was founded with the belief that the integration of big data and machine learning could revolutionize the way people travel. We gather vast amounts of up-to-date information on airfare and hotel prices. The goal of this application is to leverage this data to assist consumers in making more informed decisions when it comes to travel purchases.

In recent years, there has been an increase in air travel after the Covid-19 pandemic [3]. However, Indian airline corporations use a revenue management system to adjust prices in real time, causing fares to fluctuate considerably. This significant price fluctuation makes it essential to research price prediction. This work aims to address this issue by implementing a deep neural network system to predict flight fares.

Airlines [4] use various computational tactics to increase revenue by keeping their entire income as high as possible. These tactics include demand forecasting and pricing discrimination, which can make it difficult for customers to find the perfect value or the ideal time to purchase tickets. To estimate the amount of the flight fare, airlines use advanced computational intelligence, prediction models, and Machine Learning (ML).

6.0 Applicable Patents

US Patent 8,566,143 B2 - Performing predictive pricing based on historical data.

The Project predicts prices of items to help evaluate decisions related to them. It assists end-users in purchasing decisions and intermediate providers in selling decisions.

US Patent 8,200,514 B1 - Travel related prediction system.

The system uses historical pricing information to train a classifier for making price predictions of items. It provides price predictions for airline tickets by collecting daily flight information.

7.0 Applicable Regulations

1. **Anti-Discrimination Laws:**

- Algorithms should not lead to discriminatory outcomes. Adhere to laws that prohibit discrimination based on race, gender, age, or other protected characteristics. In the United States, the Equal Credit Opportunity Act (ECOA) is an example.

2. **Consumer Protection Laws:**

- Laws such as the Federal Trade Commission (FTC) Act and its regulations govern consumer protection. We have to be transparent about our app's functionality and ensure that it meets advertised claims.

3. **Intellectual Property Laws:**

- To abide by this law we have to respect intellectual property rights, including trademarks and copyrights. Our app should not infringe on existing patents.

4. **Algorithmic Accountability and Transparency:**

- Some jurisdictions are considering or have implemented regulations related to algorithmic transparency and accountability. We have to ensure transparency in how our algorithms make decisions, especially if it impact users significantly.

5. **Electronic Communications Privacy Act (ECPA):**

- Regulates interception of electronic communications. Our app should comply with rules related to wiretapping and electronic surveillance.

8.0 Applicable Constraints

When developing a web app that incorporates machine learning algorithms, various constraints need to be considered to ensure the app's effectiveness, ethical use, and compliance with legal and technical requirements. Here are some applicable constraints:

1. **Data Privacy and Security:**

- Ensuring that user data is handled securely, and implementing measures to protect against unauthorized access, data breaches, or other security threats.

2. **Interpretability and Explainability:**

- Constraints related to the interpretability and explainability of machine learning models. Depending on the application, it may be crucial to provide users with insights into how the algorithm makes decisions.

3. **Resource Limitations:**

- Constraints on computational resources, especially if the web app needs to handle a large volume of data or real-time predictions. This optimizes algorithms for efficiency and scalability to meet resource constraints.

4. **User Experience:**

- We have to keep in mind that the machine learning features enhance, rather than hinder, the overall user experience. Strive for simplicity and clarity in presenting predictions or recommendations.

5. **Data Quality and Availability:**

- Constraints on data quality and availability can impact the performance of machine learning models. The training data we collect should be of high quality so the model can handle variations in real-world data.

9.0 Business Model

Airways are one of the fastest modes of transportation but the prices are not customer friendly. Our proposed web application can be a solution to these problems. The model's prediction will enable customers to secure reasonably priced tickets. Travel agencies operating on a smaller scale can utilize this web application to facilitate ticket bookings for their clients.

By partnering with travel agencies, we could earn commissions for each successful flight booking made through your web application. The system will feature affiliate links when users view forecasted prices. Travel agencies can benefit from providing the web app as a service to their customers. The web app will earn commissions for directing users to platforms for booking and establishing a mutually beneficial revenue stream. The focus on user-centric features and continuous innovation ensures the app's competitiveness in the dynamic air travel market. Using this model we can forge partnerships with small travel agencies in the local area.

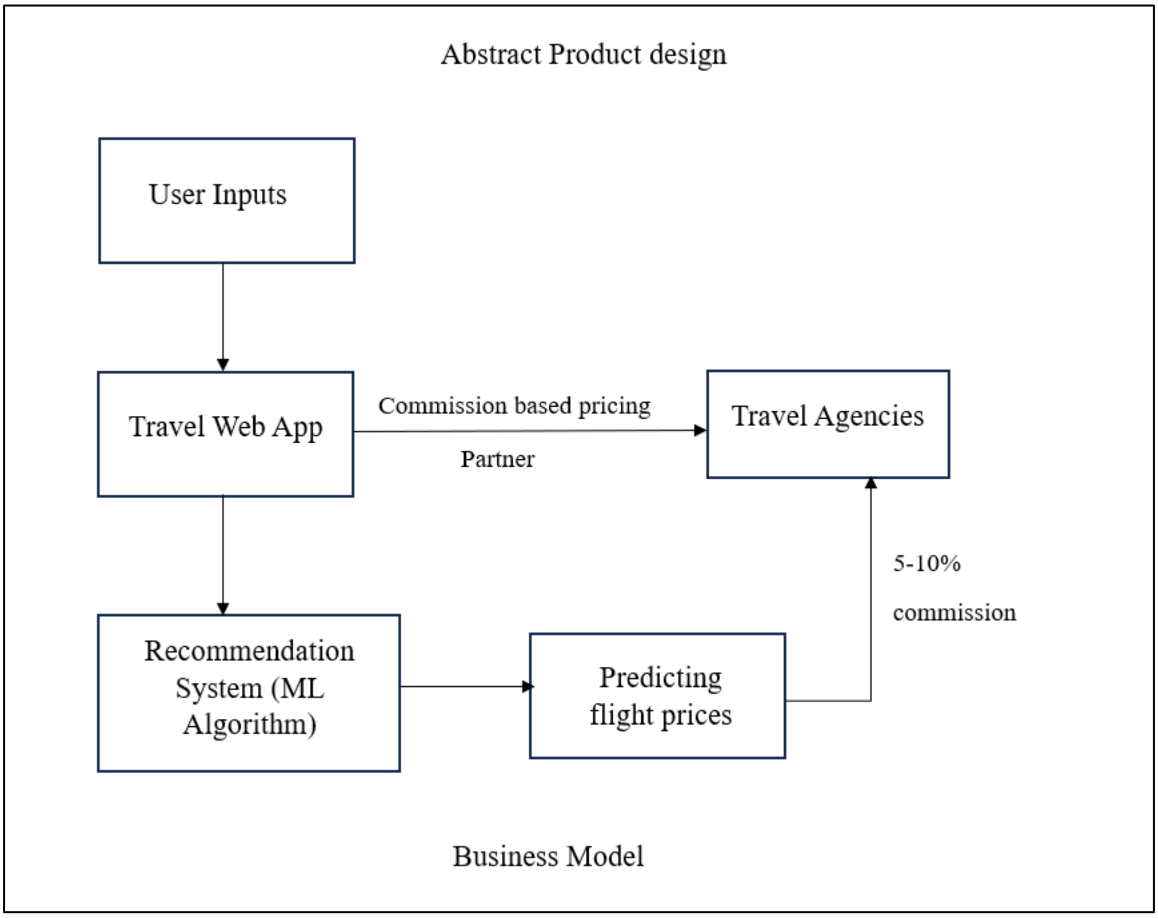

Figure 2: Abstract Product Design

10. Concept Generation

Customers find it difficult to find the right value and the ideal time to purchase tickets. Machine Learning Algorithms play an important role in this sector as they study historical data of flight tickets and their trends to predict accurate results. The ability to scrape flight prices from airlines is nothing short of a game changer for providing travellers with the best deals available in the ever-changing world of air travel. This data can accurately forecast prices with the right Machine Learning model.

During the booking process, the customers won't be confused as the model will recommend them to wait if there is an upcoming downtrend in the prices and will also provide an approximate date when the flight ticket will be the cheapest. The presentation will also display an estimate of the amount saved through the use of the algorithm.

11. Concept Development

The data for the proposed web application can be obtained by airlines companies and then cleaned and transformed using feature engineering to prepare it for algorithm creation. The data is then used for training and testing the model. For model development, we will be using algorithms such as Decision Tree Regressor, Random Forest and Gradient Boosting Regressor. VS Code or any other IDE can be used for the code implementation. After the development process, the model with good accuracy, will be pushed to GitHub. The web application will be developed using Flask and the machine learning model, and then deployed to a suitable cloud platform.

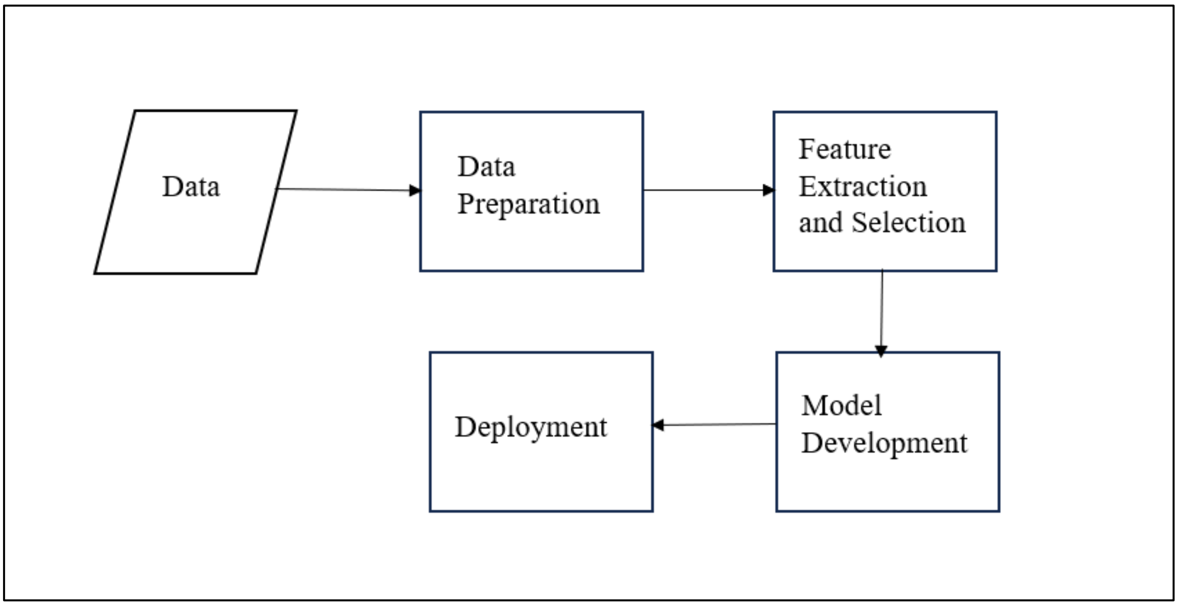

Figure 3: Concept Development

12. Final Product

The final product will have the following segments:

12.1. Back End

Data Preparation

The proposed system will require a huge amount of data to have a competitive prediction score. The data will be cleaned and transformed using libraries such as Pandas and NumPy. To collect the data continuously we would have to automate this manipulation process using Python.

Model Creation and Evaluation

The manipulated data is used to train the data. Using the right algorithm can give rise best results. Supervised machine learning algorithms such as Decision Tree Regressor, Random Forest and Gradient Boosting Regressor will be used in the prediction process. The model with the best accuracy will be deployed to the web application.

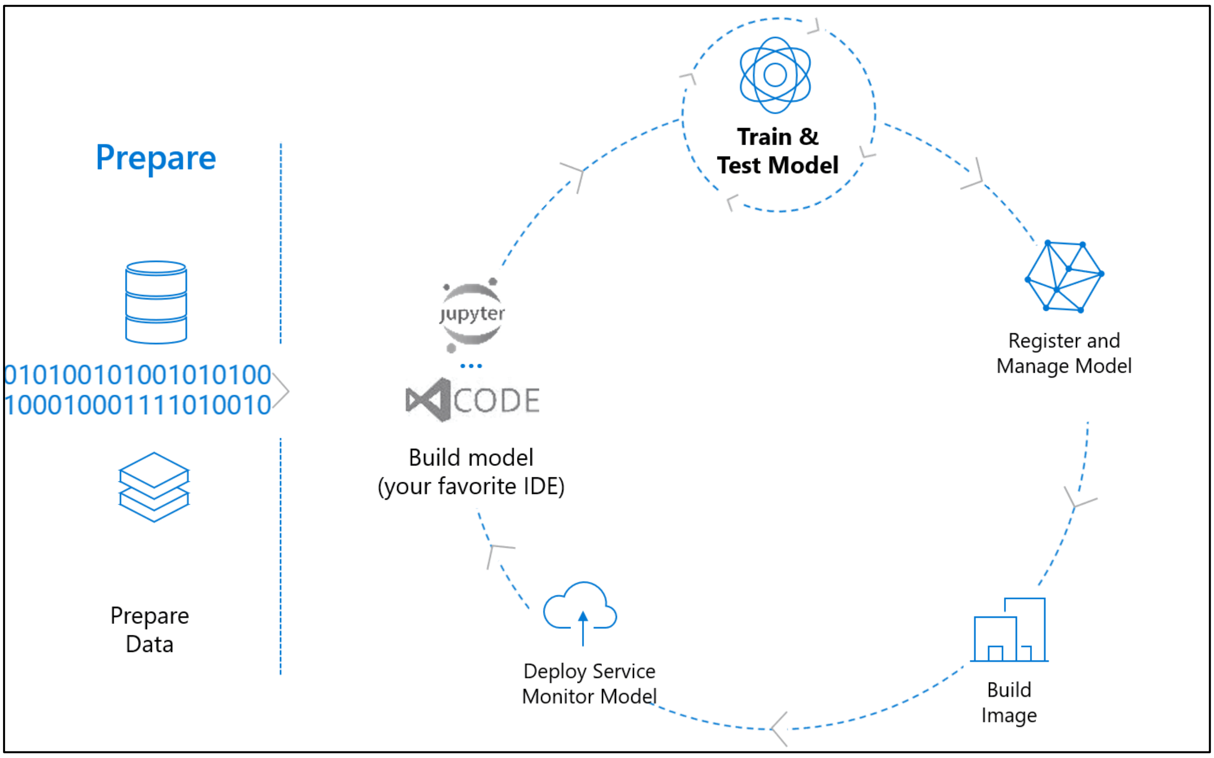

Figure 4: Model Creation Cycle

12.2. Front End

Users input their preferences just like they do in a normal travel app. However, after entering their data and choosing the flight, the algorithm will present the forecast about that particular flight booking. The users will be provided with data with the help of visualization techniques. The approximate date on when the ticket will be the cheapest will also be provided so that the users can make their decision.

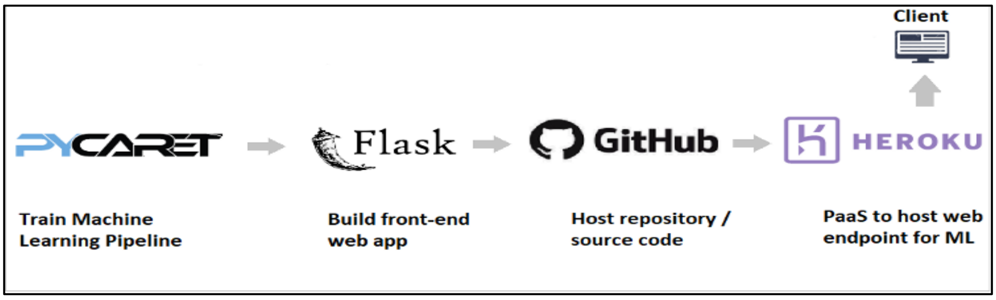

Figure 5: Abstract Web App Design

13. Product Details

13.1 Algorithms

Decision Tree

The decision tree is an algorithm that uses tree like structure that includes all the inputs, and the possible outcomes. Decision tree falls under the category of supervised learning and it is used for both continuous as well as categorical outputs. There are two nodes one is the Decision node and the other is the Leaf node. The Decision node is used to make decisions whereas the Leaf node is the output of these decisions and there are no further nodes.

Decision trees try to develop the ability to imitate human thinking when making decisions. The algorithm starts from the root node and compares the values of the root attribute with the real dataset attribute and goes on to the next node. The model follows the same algorithm and continues the process until it reaches the leaf node.

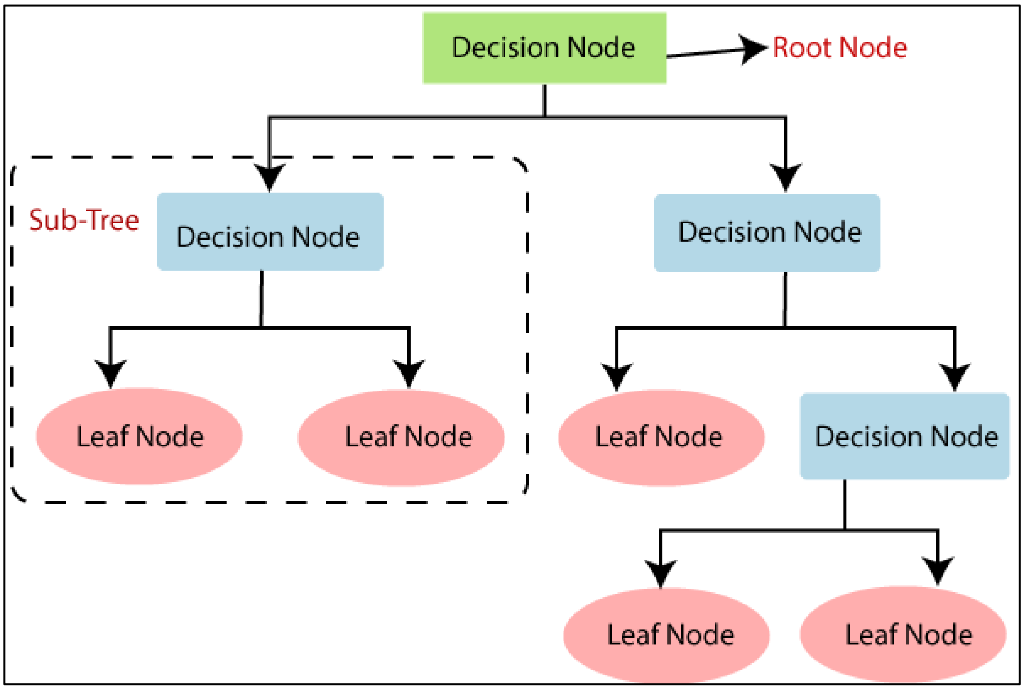

Figure 6: Decision Tree

Random Forest

Random forest is an ensemble learning that combines multiple classifiers to improve the performance of the model. It contains several decision trees based on different subsets of the dataset. The algorithm takes the average of all the outputs of the decision trees and predicts the final output to improve the accuracy of the model.

In this algorithm, the greater the number of trees the higher the accuracy of the model and it also prevents the problem of overfitting. Due to this, the model prediction for large datasets is highly accurate and it does this with less training time compared to other algorithms.

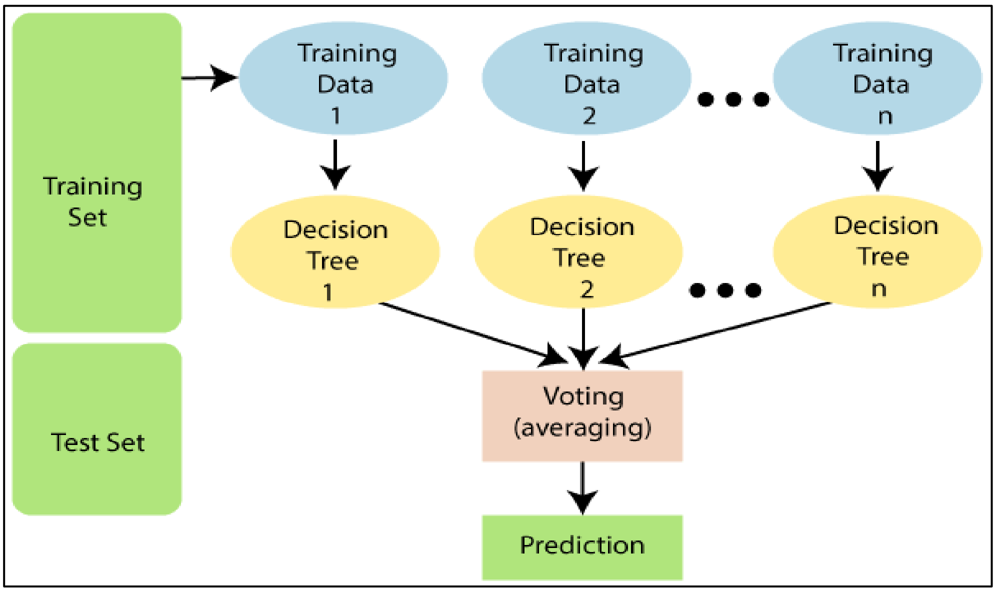

Figure 7: Random Forest

Gradient Boosting

Gradient Boosting is a machine learning algorithm that is used for classification and regression problems. Gradient Boosting is a type of ensemble learning which trains numerous models at a time and then it tries to rectify the mistake of the previous model. Each model is trained to minimize the loss function of the previous model. In each iteration, the algorithm calculates the gradient of the loss function and the current model's prediction and then trains a new weak model. This process is repeated further until the conditions are met.

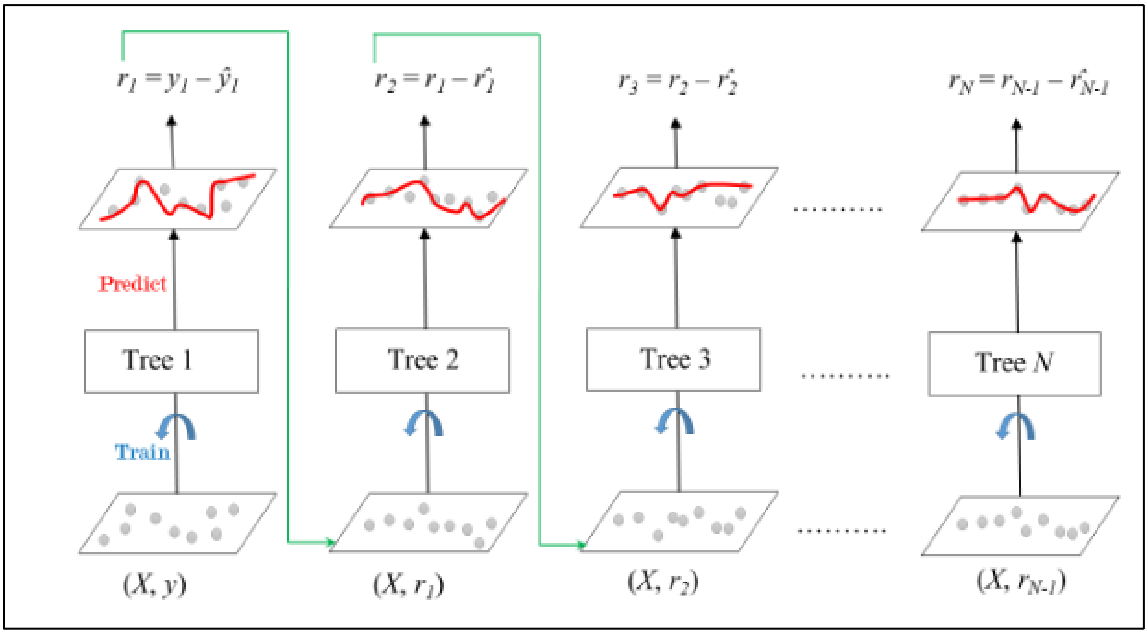

Figure 8: Gradient Boosting

14. Code Implementation

14.1. Exploratory Data Analysis

For the purpose of this report, we will be using a Kaggle dataset which will be used to train and test the model.

Here is the dataset link: https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction/data

[4]:

```
df = pd.read_csv("/kaggle/input/flight-price-prediction/Clean_Dataset.csv")
df.head(10)
```

[4]:

	Unnamed: 0	airline	flight	source_city	departure_time	stops	arrival_time	destination_city	class	duration	days_left	price
0	0	SpiceJet	SG-8709	Delhi	Evening	zero	Night	Mumbai	Economy	2.17	1	5953
1	1	SpiceJet	SG-8157	Delhi	Early_Morning	zero	Morning	Mumbai	Economy	2.33	1	5953
2	2	AirAsia	I5-764	Delhi	Early_Morning	zero	Early_Morning	Mumbai	Economy	2.17	1	5956
3	3	Vistara	UK-995	Delhi	Morning	zero	Afternoon	Mumbai	Economy	2.25	1	5955
4	4	Vistara	UK-963	Delhi	Morning	zero	Morning	Mumbai	Economy	2.33	1	5955
5	5	Vistara	UK-945	Delhi	Morning	zero	Afternoon	Mumbai	Economy	2.33	1	5955
6	6	Vistara	UK-927	Delhi	Morning	zero	Morning	Mumbai	Economy	2.08	1	6060
7	7	Vistara	UK-951	Delhi	Afternoon	zero	Evening	Mumbai	Economy	2.17	1	6060
8	8	GO_FIRST	G8-334	Delhi	Early_Morning	zero	Morning	Mumbai	Economy	2.17	1	5954
9	9	GO_FIRST	G8-336	Delhi	Afternoon	zero	Evening	Mumbai	Economy	2.25	1	5954

Figure 9: Dataset

In the below graph, we compare the Airline Companies and their Prices to determine how it affect the ticket prices.

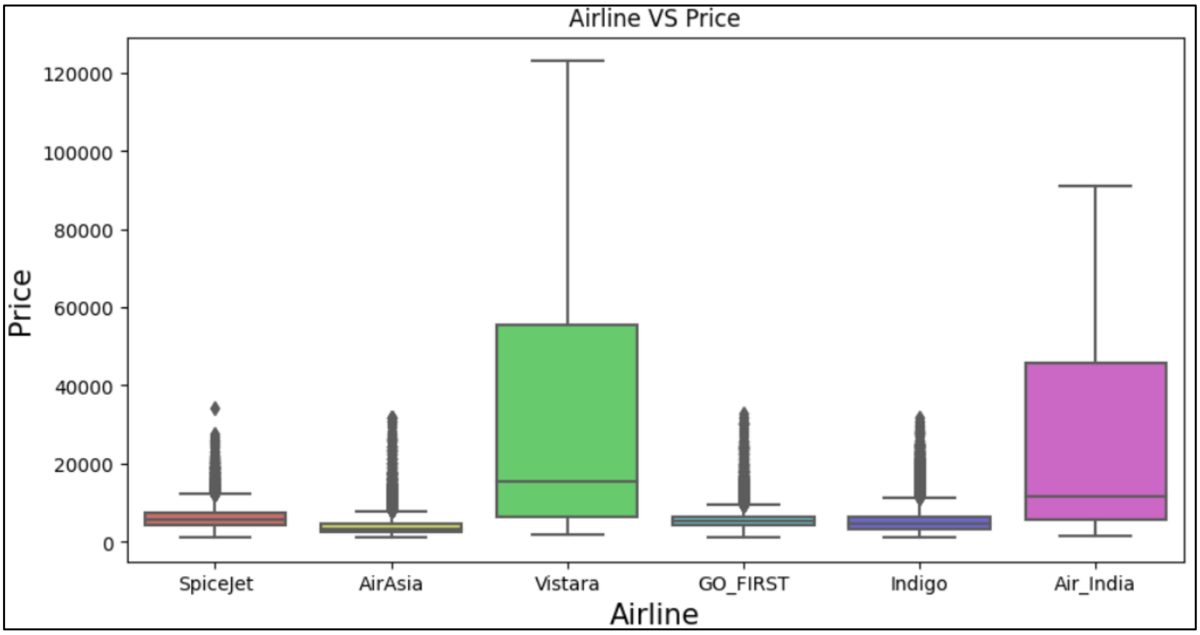

Figure 10: Price vs Airline

The graph below displays a comparison of prices for flight tickets across different travel classes.

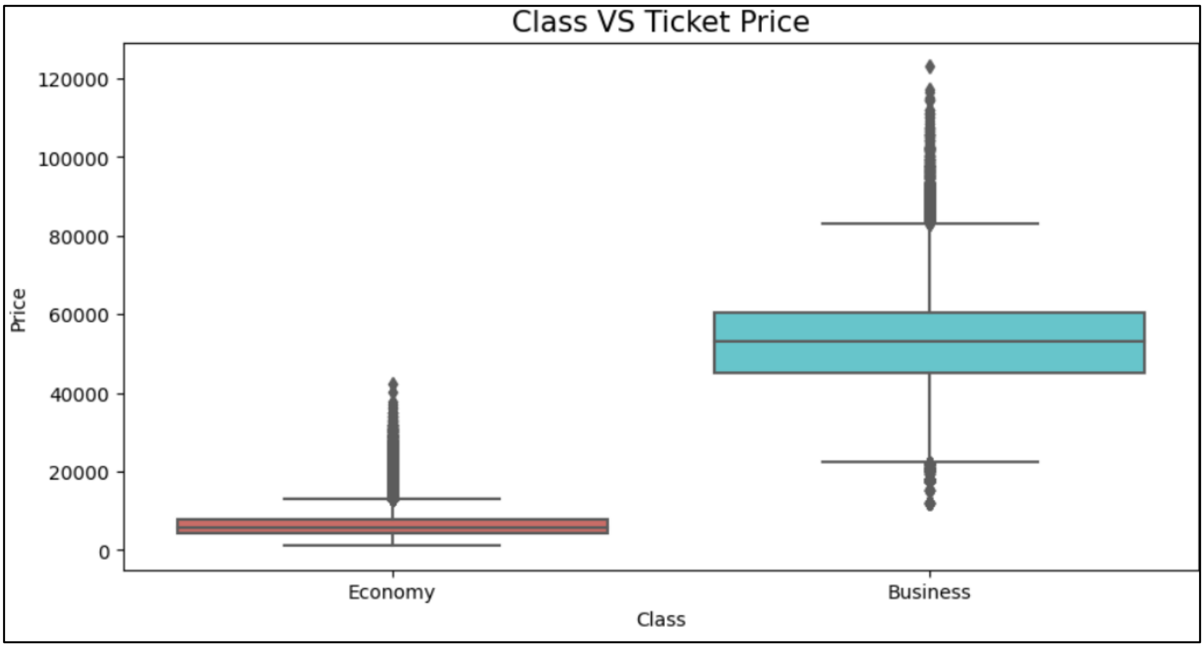

Figure 11: Class vs Price

Here is a comparison of how the number of stops in a flight affects the price.

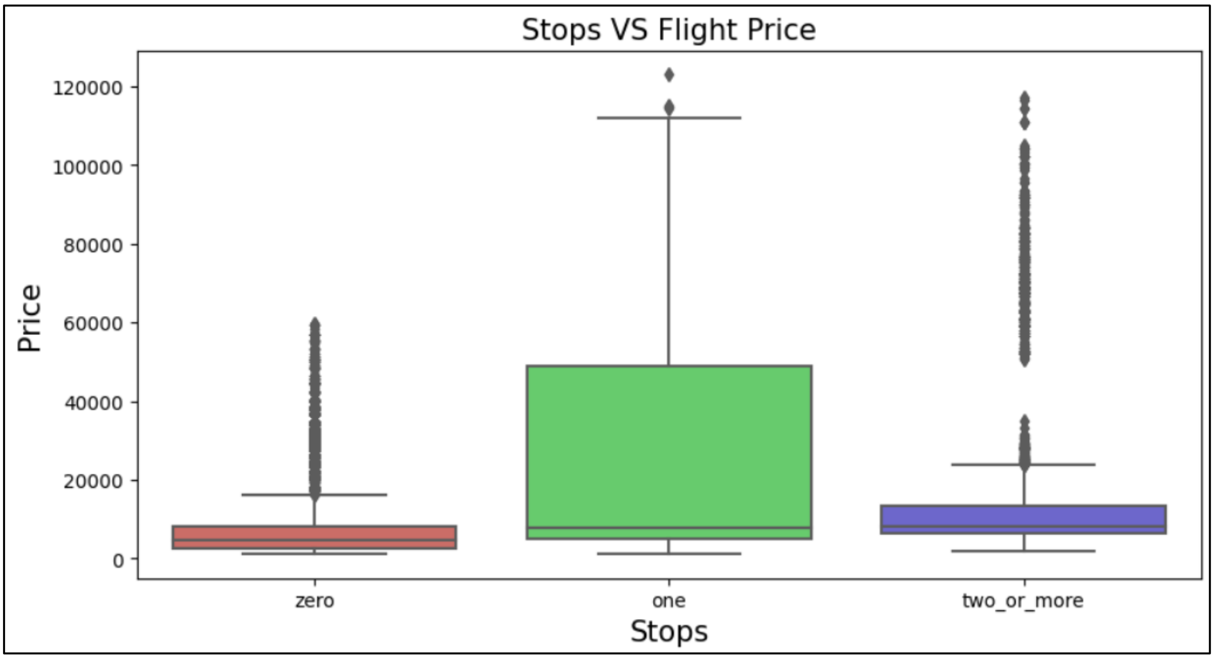

Figure 12: Number of Stops vs Price

The line graph shows how the number of departure days affects airline tickets.

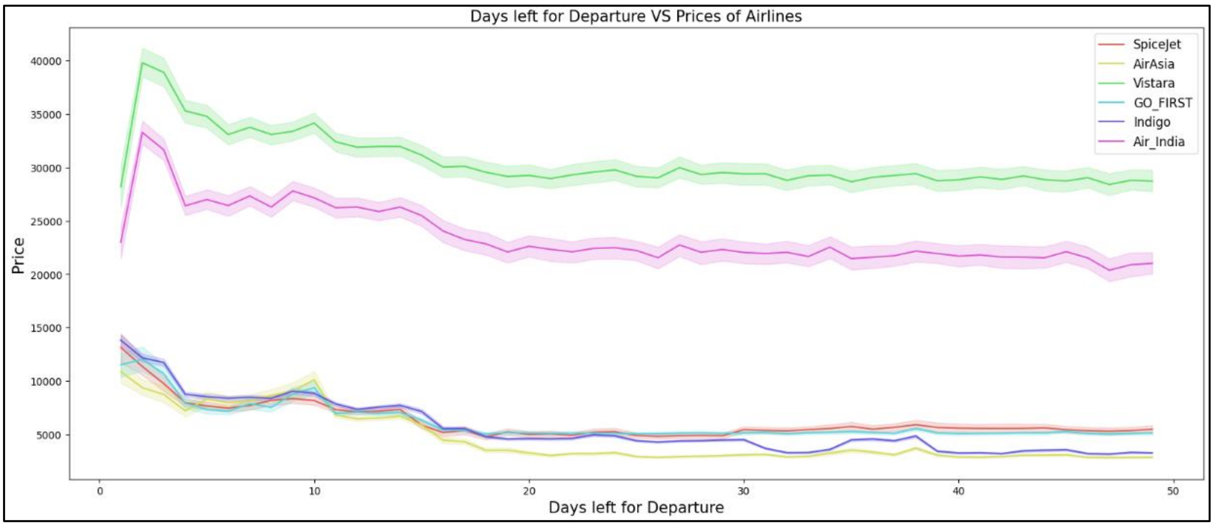

Figure 13: Departure Days vs Price

14.2. ML Modelling

During the model building phase, we used the Decision Tree Regressor, Random Forest Regressor and Gradient Boosting algorithm.

Decision Tree Regressor

```
dt = DecisionTreeRegressor(random_state=42)

dt.fit(x_train, y_train)

# Predict the values for the training and testing sets
y_pred_train = dt.predict(x_train)
y_pred_test = dt.predict(x_test)

# Compute the accuracy, MSE, and R2 for the training set
accuracy_train = dt.score(x_train, y_train)
mse_train = mean_squared_error(y_train, y_pred_train)
r2_train = r2_score(y_train, y_pred_train)

# Compute the accuracy, MSE, and R2 for the testing set
accuracy_test = dt.score(x_test, y_test)
mse_test = mean_squared_error(y_test, y_pred_test)
r2_test = r2_score(y_test, y_pred_test)

print("Accuracy - Train: {:} Test: {:}".format(accuracy_train, accuracy_test))
print("MSE - Train: {:} Test: {:}".format(mse_train, mse_test))
print("R2 - Train: {:} Test: {:}".format(r2_train, r2_test))
```

```
Accuracy - Train: 0.9993813250067048 Test: 0.9461809304968603
MSE - Train: 318871.17609519593 Test: 27698872.227363735
R2 - Train: 0.9993813250067048 Test: 0.9461809304968603
```

Figure 14: Decision Tree

Random Forest Regressor

```
rf = RandomForestRegressor(n_estimators=10, random_state=42)

rf.fit(x_train, y_train)

# Predict the values for the training and testing sets
y_pred_train = rf.predict(x_train)
y_pred_test = rf.predict(x_test)

# Compute the accuracy, MSE, and R2 for the training set
accuracy_train = rf.score(x_train, y_train)
mse_train = mean_squared_error(y_train, y_pred_train)
r2_train = r2_score(y_train, y_pred_train)

# Compute the accuracy, MSE, and R2 for the testing set
accuracy_test = rf.score(x_test, y_test)
mse_test = mean_squared_error(y_test, y_pred_test)
r2_test = r2_score(y_test, y_pred_test)

print("Accuracy - Train: {:} Test: {:}".format(accuracy_train, accuracy_test))
print("MSE - Train: {:} Test: {:}".format(mse_train, mse_test))
print("R2 - Train: {:} Test: {:}".format(r2_train, r2_test))
```

```
Accuracy - Train: 0.9968919955333009 Test: 0.9624071138728673
MSE - Train: 1601896.0687693392 Test: 19347799.192859273
R2 - Train: 0.9968919955333009 Test: 0.9624071138728673
```

Figure 15: Random Forest

Gradient Boosting Regressor

```
gb = GradientBoostingRegressor(loss='squared_error', learning_rate=0.1, n_estimators=100, max_depth=3)

gb.fit(x_train, y_train)

y_pred_train = gb.predict(x_train)
y_pred_test = gb.predict(x_test)

accuracy_train = gb.score(x_train, y_train)
mse_train = mean_squared_error(y_train, y_pred_train)
r2_train = r2_score(y_train, y_pred_train)


accuracy_test = gb.score(x_test, y_test)
mse_test = mean_squared_error(y_test, y_pred_test)
r2_test = r2_score(y_test, y_pred_test)

print("Accuracy - Train: {:} Test: {:}".format(accuracy_train, accuracy_test))
print("MSE - Train: {:} Test: {:}".format(mse_train, mse_test))
print("R2 - Train: {:} Test: {:}".format(r2_train, r2_test))
```

```
Accuracy - Train: 0.9523303311529738 Test: 0.9514550798583008
MSE - Train: 24569416.145881303 Test: 24984444.22592734
R2 - Train: 0.9523303311529738 Test: 0.9514550798583008
```

Figure 16: Gradient Boosting

Accuracy Table

No.	Algorithm	Accuracy
1	Decision Tree Regressor	0.9461
2	Random Forest Regressor	0.9624
3	Gradient Boosting Regressor	0.9514

14.3. GitHub Link

The GitHub link of the code: https://github.com/Spencerdsheel/airline-price-prediction

15. Conclusion

In conclusion, the development of our travel project represents a stride towards revolutionizing the industry with the help of machine learning. With a meticulous business needs assessment, we are certain our web application would make informed decisions and enhance customer experience. Our innovative approach opens doors to further possibilities.

In essence, this project is a gateway to a more seamless and personalized travel experience. We remain steadfast in our commitment to excellence, innovation and the evolving needs of the modern traveller.

16. References

[1] K. T. N. K. A. N. E. V. a. G. A. P. T. Kalampokas, “A Holistic Approach on Airfare Price Prediction Using Machine Learning Techniques,” IEEE Access, 2023.

[2] N. S. S. V. S. R. a. S. J. J. Thangaraj, “Flight Ticket Prediction using Random Forest Regressor Compared with Decision Tree Regressor,” 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023.

[3] G. Ratnakanth, “Prediction of Flight Fare using Deep Learning Techniques,” International Conference on Computing, Communication and Power Technology (IC3P), Visakhapatnam, India, 2022.

[4] N. Alapati, “Prediction of Flight-fare using machine learning,” International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP), Uttarakhand, India, 2022.