ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

DE VINCI
EXECUTIVE
EDUCATION

ECO-FRIENDLY FLIGHT PREDICTIONS: BALANCING COST AND SUSTAINABILITY WITH MACHINE LEARNING

Supervised by:
Dr.Nédra MELLOULI

Presented by:
Anis HENTIT Hamza HALINE

Machine learning For CV

TABLE OF CONTENTS

01 Introduction - Aim & Hypothesis & Problem Definition

02 Data Collection and Processing

03 Methodology used

04 Problems faced and Solutions

05 Demonstration of our streamlit app

06 Conclusion

INTRODUCTION

Aim:

Build an interactive tool that predicts flight prices and integrates ecological metrics like carbon emissions, promoting affordable and sustainable travel.

Hypothesis:

- Predicting flight prices and integrating environmental metrics (like emissions) can help balance cost and sustainability.
- Travelers are likely to opt for eco-friendly flights if provided with clear metrics.

PROBLEM DEFINITION

- Travelers often lack tools to balance cost and ecological impact when booking flights.
- Current platforms do not integrate price prediction with sustainability metrics (e.g., carbon emissions).

Goal:

Develop a system that:
Predicts flight prices accuratelyand and Provides insights into ecological impact using a custom EcoScore metric.

DATA COLLECTION

Kayak Platform

-Used Selenium WebDriver to scrape flight data specifically for routes between Paris airports (CDG and ORY) and 20 international destinations over a period of 90 days.
-The data captured from Kayak included:
-Ticket prices, number of stops, flight duration, airline details, and seat types.

Amadeus API:

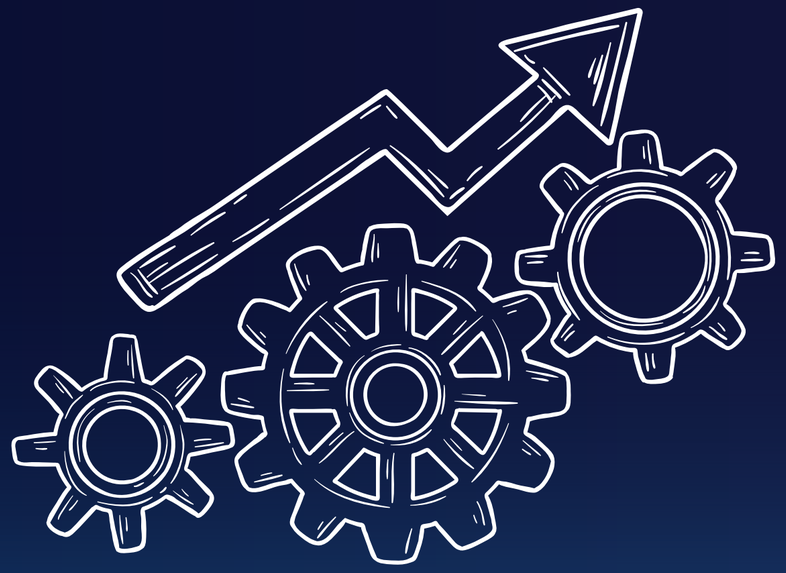
-We leveraged the Amadeus API to scrape over 180,000 flight records from destinations worldwide.

-The data included a wide range of features like flight duration, number of stops, ticket prices, and airline details.

DATA PROCESSING

Data Cleaning:

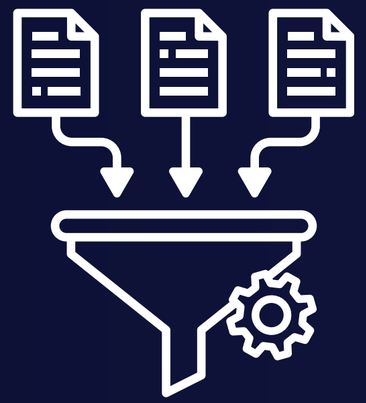
- Removed duplicate flight records to ensure unique data entries.
- Standardized key features:
  - Stops: Converted text (e.g., "1 stop") into numeric values.
  - Price: Stripped currency symbols and converted to integers.
  - Time: Unified departure and arrival times into a consistent form



Feature Engineering:

Grouped Seat Types: Consolidated into "Economy," "Business," etc.
Multi-Carrier Indicator: Identified flights involving multiple airlines.
Distance Calculation: Used geodesic coordinates of airports.
Carbon Emissions:Computed as:
Emissions (kg CO_2) = Distance (km) × 0.115 + Stops × 50

DATA PROCESSING



Dataset Integration:

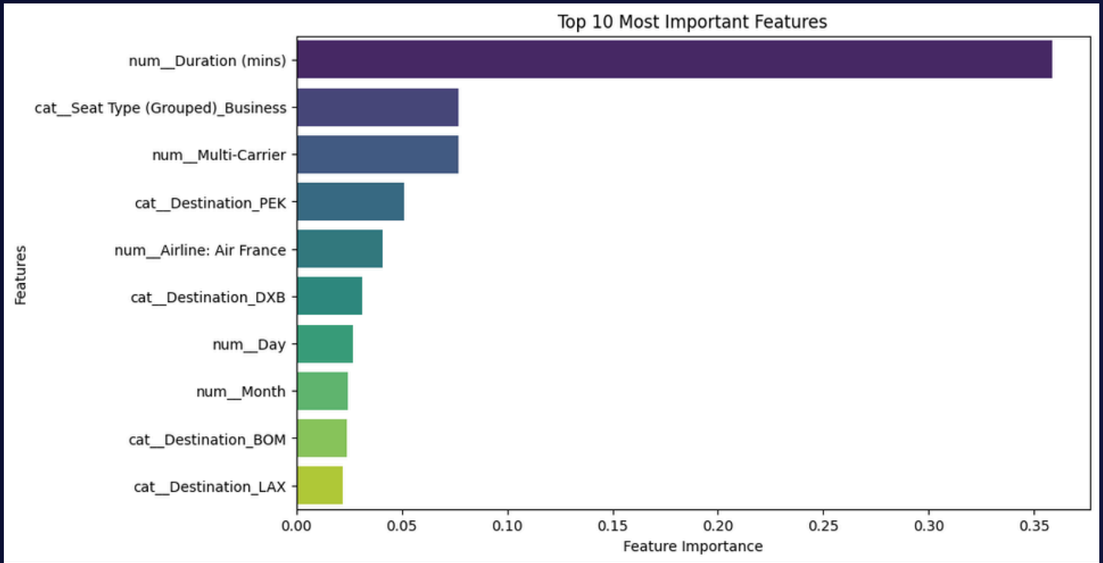
- Combined data from Kayak and Amadeus API into a unified structure.
- Retained only essential columns for analysis and modeling.

Final Outcome:

The dataset included both categorical (origin, destination, airlines) and numerical (stops, distance, emissions) features, ready for exploratory analysis and machine learning.

METHODOLOGY USED

Exploratory Data Analysis (EDA):



Identified relationships between key features:

- **Strong correlation between price and duration (0.46).**
- **Moderate correlation between price and stops (0.36).**

XGBoost Model:

Trained two models:

- **Price Prediction Model: Focused solely on predicting flight prices.**
- **Eco-Friendly Model: Optimized for the custom EcoScore metric:**

$$\text{EcoScore} = \alpha \cdot \text{Price Error} + \beta \cdot \text{Carbon Emissions}$$

```
Fitting 3 folds for each of 48 candidates, totalling 144 fits
c:\Users\Anis\AppData\Local\Programs\Python\Python311\Lib\site-package
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan n
 nan nan nan nan nan nan nan nan nan nan nan]
  warnings.warn(

XGBoost with EcoScore - Training Performance:
MAE: 26.92, RMSE: 73.32, R²: 0.94

XGBoost with EcoScore - Testing Performance:
MAE: 27.16, RMSE: 75.53, R²: 0.93
```

PROBLEMS AND SOLUTIONS

Model Used: YOLO11

- Pre-trained weights: last.pt (if available, resuming training).
- Lightweight model: yolo11n.pt (used if starting fresh, optimized for T4 GPUs).

Model Used: YOLO11

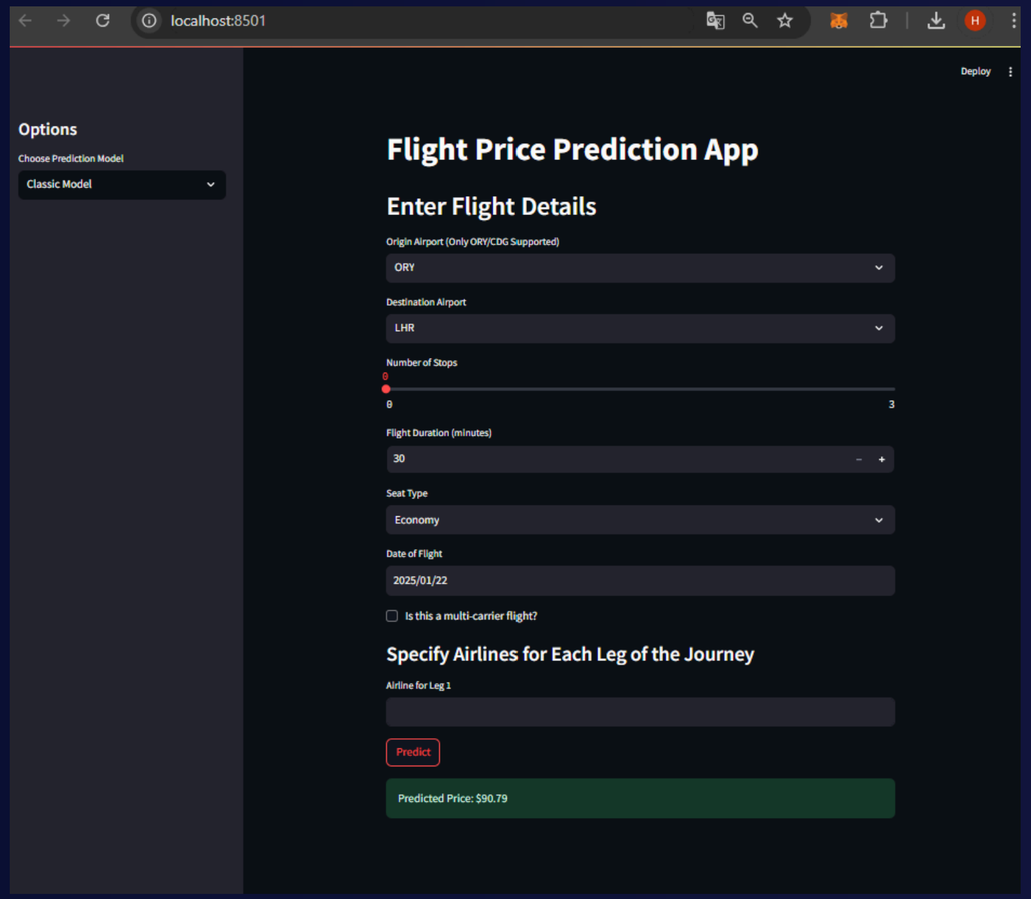
- Pre-trained weights: last.pt (if available, resuming training).
- Lightweight model: yolo11n.pt (used if starting fresh, optimized for T4 GPUs).

PROBLEMS AND SOLUTIONS

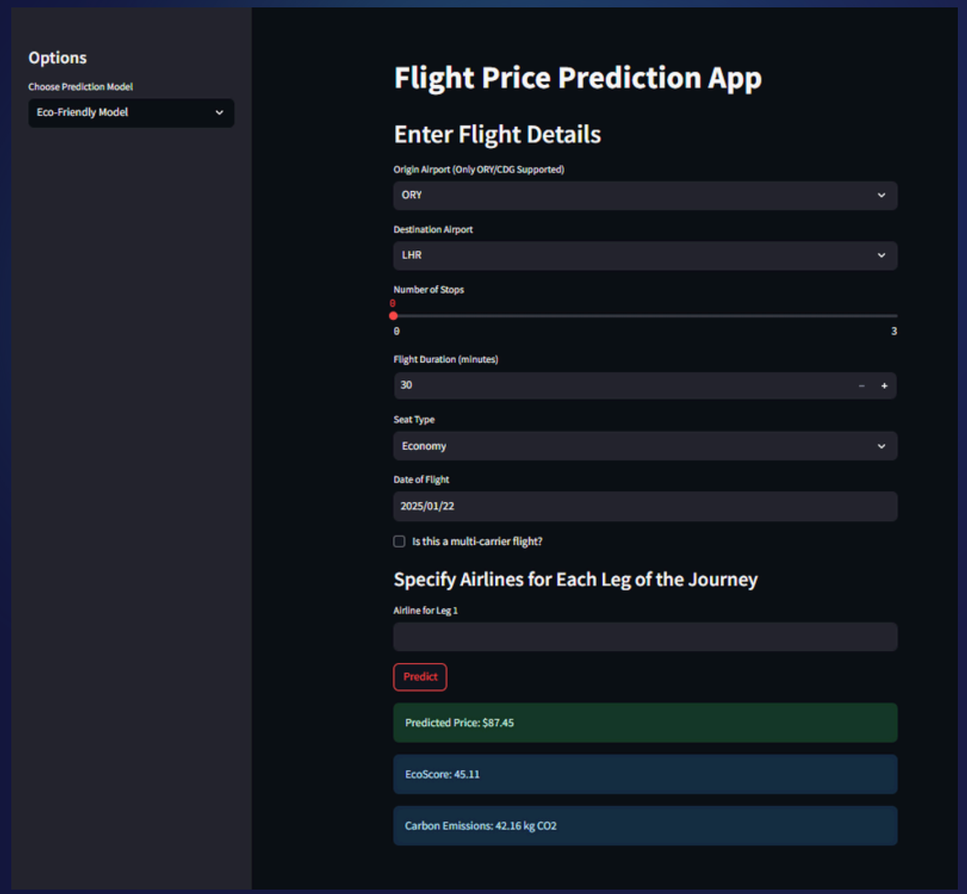
Dynamic Content Loading	Implemented incremental scrolling and adaptive element selection using Selenium.
CAPTCHAs	Rotated user agents and added random delays to mimic human behavior and bypass detection.
Balancing Cost and Ecology	Developed a custom EcoScore metric that balances price prediction with carbon emissions.
API Data Collection	Managed data collection in smaller batches and used multiple keys to maximize coverage while reducing downtime.

DEMONSTRATION OF OUR STREAMLIT APP

Classic Model



Eco-Friendly Model



CONCLUSION

This project successfully combined machine learning and ecological considerations to create a user-friendly Streamlit app that predicts flight prices and evaluates environmental impact using the EcoScore metric. It highlights the importance of balancing affordability and sustainability in travel decisions. Future improvements include integrating real-time data, expanding destination coverage, and refining the EcoScore to promote responsible tourism further.

THANK YOU

ANY QUESTIONS?