IPSW 2024 Air Canada Final Presentation

Dynamic Air Cargo Optimization Using Machine Learning Models

Georgios Farfaras, Flora Gao, Vasanth Ramkumar, Janik Gagne, Chad Saad (Air Canada)

Sylvain Perron, Ernest Tafolong, Golshid Aflaki (HEC)

Faramarz Farhangian (ETS)

Agenda

- Problem Description
- Data Description
- Methodology
- Results
- Conclusion and Future Works

Problem Description

Manual optimization of spot rates is **inefficient** and can lead to **suboptimal revenue**, requiring an automated solution to enhance pricing strategies.

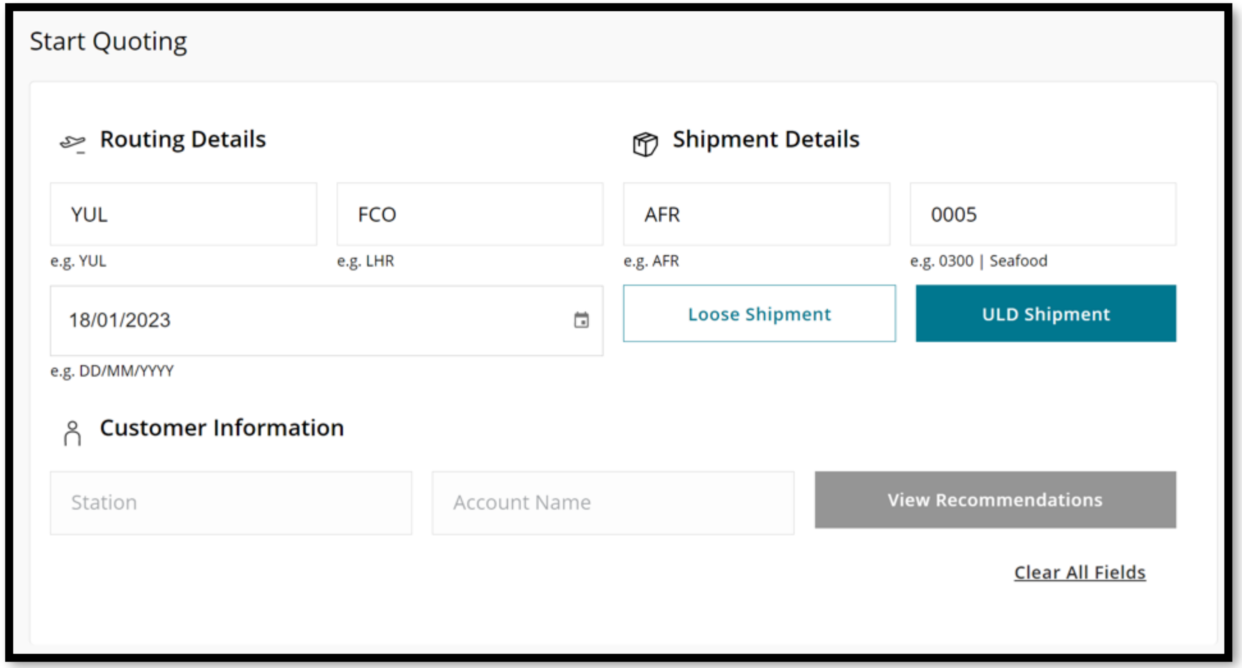
Goal of the Project

Develop a machine learning model to quickly test and optimize spot rate factors, aiming to **maximize revenue** with consistent and explainable recommendations.

Vector | AI
AIR CANADA CARGO

Data Description

1. Customers call Air Canada for a shipment quote.
2. Agents collect **shipment**, **route**, and **account details**
3. Provide the best rate per kg based on **<u>contract rates, flight load factors, Density</u>** and **<u>days to departure.</u>**



Data Description

We have historical data (More than 150,000 records) from **2018** to **2024** on **shipment details**, **account details**, **route details**, **recommended rates, current load factor**, **days to departure**, **chargeable weight**, and **quote status** (whether quotes were accepted or rejected by customers).

Challenges

- **Missing Values:** Incomplete data entries that need to be addressed to ensure model accuracy.
- **Outliers:** Extreme values that can skew results and need to be identified and managed.
- **Pandemic Data (2020-2022):** Data from this period may be atypical and should be excluded to avoid distortions.
- **Imbalanced Data:** Disproportionate number of accepted versus rejected quotes, complicating model training.
- **Data Unavailability:** Lack of access to crucial data such as market trends, impacting model comprehensiveness.
- **Mixed Markets:** Different market data are combined, complicating analysis and model accuracy.
- **Post-2019 Changes:** Significant shifts since 2019 that affect the relevance of older data.
- **Missing Important Features:** The absence of key features like marginal cost, adjustment load factor, base rate, density adjustment factor, and expected load factor, which are crucial for accurate rate recommendations.

Methodology

Revenue Optimization Simulation

1. **Simulate Requests**: Simulate a set of requests for one flight (Bootstrap resampling)
2. **Generate Quoted Rates**: Apply an adjustment factors matrix to generate quoted rates for each simulated request.
3. **Calculate Acceptance Probability**: Use a logistic regression (Random Forest) model to calculate the probability of each quote being accepted.
4. **Compute Total Revenue**: Calculate the total revenue from all accepted quotes.
5. **Iterate and Compare**: Repeat steps 2-4 with different adjustment matrices and compare the resulting revenues to find the optimal rates.

Simulate Requests

First, we generate a subset of shipment requests based on specific settings:

- **Market:** North America to Europe
- **Dates:** July 2023 to April 2024

For each simulated data point, we include the following information:

- **Days Out:** Days before departure (negative value)
- **Quoted Rate:** The rate quoted to the customer
- **Base Rate:** The initial rate before adjustments
- **Quote Status: 1** if accepted by the client; **0** otherwise
- **Chargeable Weight:** The weight of the shipment
- **Current Load Factor:** The load factor of the flight at the time of the quote

Generate Quoted Rates

Vector | AI
AIR CANADA CARGO

Expected LF: OPEN							
Days to Departure / Current LF	**0 to 1**	**1 to 2**	**2 to 3**	**3 to 4**	**4 to 5**	**5 to 10**	**10 to 15**
0% - 30%	0.6	0.6	0.6	0.6	0.75	1	1
30% - 45%	0.6	0.6	0.6	0.6	0.85	1	1
45% - 60%	0.7	0.7	0.79872	0.79904	0.95	1	1
60% - 75%	1.1	1.018433	1.004747	1	1	1	1
75% - 90%	1.3	1.238467	1.210773	1.18308	1.155387	1.127693	1.05
90% - 100%	1.5	1.4585	1.4168	1.3751	1.3334	1.2917	1.05
>100%	2.74	2.575	2.41	2.245	2.08	1.915	1.75
Expected LF: AVERAGE							
Days to Departure / Current LF	**0 to 1**	**1 to 2**	**2 to 3**	**3 to 4**	**4 to 5**	**5 to 10**	**10 to 15**
0% - 30%	0.7	0.7	0.7	0.7	0.85	1	1
30% - 45%	0.7	0.7	0.7	0.7	0.9	1	1
45% - 60%	0.75	0.75	0.84872	0.87404	1	1	1
60% - 75%	1.1	1.04404	1.04106	1.03808	1.0351	1	1
75% - 90%	1.3	1.2468	1.22744	1.20808	1.18872	1.16936	1.1
90% - 100%	1.5	1.4585	1.4168	1.3751	1.3334	1.2917	1.1
>100%	2.74	2.58	2.41	2.25	2.08	1.92	1.75
Expected LF: CONSTRAINED							
Days to Departure / Current LF	**0 to 1**	**1 to 2**	**2 to 3**	**3 to 4**	**4 to 5**	**5 to 10**	**10 to 15**
0% - 30%	0.8	0.8	0.8	0.8	1	1	1
30% - 45%	0.8	0.8	0.8	0.8	1	1	1
45% - 60%	0.8	0.8	0.88208	0.92406	1	1	1
60% - 75%	1.1	1.043467	1.054773	1.06608	1.077387	1.05	1.1
75% - 90%	1.3	1.246833	1.227467	1.2081	1.188733	1.169367	1.1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

We use an adjustment factors matrix including:

- **Current Load Factor**
- **Days to Departure**

To generate quoted rates for the simulated requests.

Version Feb 2024

Calculate Acceptance Probability

We employ a Logistic Regression Model to predict the probability of quote acceptance. Logistic regression estimates the probability of quote acceptance based on these variables.

- **Days Out:** Negative value representing days before departure
- **Quote Status**: 1 for accepted, 0 for rejected
- **Chargeable Weight**
- **Current CAD Rate**

Calculate Acceptance Probability

- Current CAD rate
- Chargeable Weight
- Interaction of days left to departure and current rate.

- Day left to departure
- Interaction of days to departure and the current rate

```
Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              1.021e+00  3.977e-02  25.669  < 2e-16 ***
days_out                -7.734e-03  4.024e-03  -1.922   0.0546 .
cad_rate                -5.732e-01  2.442e-02 -23.470  < 2e-16 ***
chargeable_wgt          -2.117e-04  3.350e-05  -6.320 2.62e-10 ***
cad_rate:chargeable_wgt -3.755e-05  2.167e-05  -1.733   0.0830 .
days_out:cad_rate        5.570e-03  2.486e-03   2.241   0.0250 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Calculate Acceptance Probability

Logistic Regression Predictor

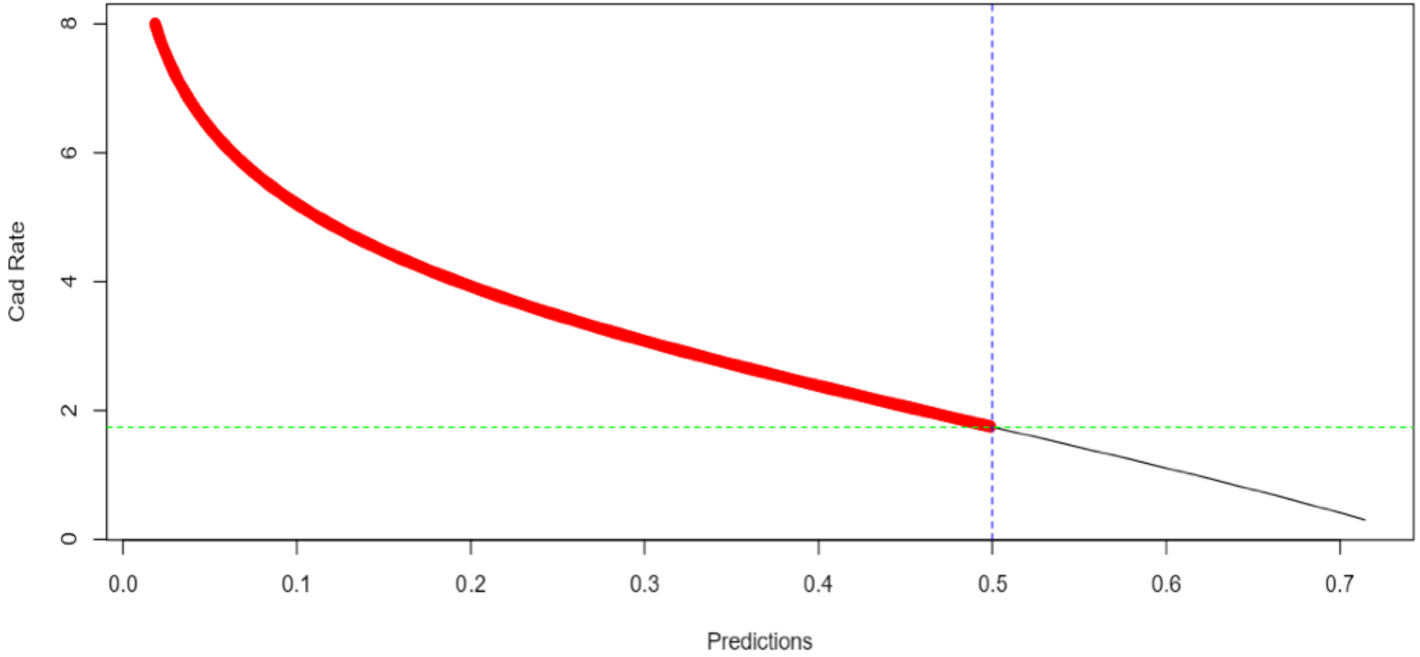
Days Out:

-5

Weight:

10

Predict



Optimization Step

Compute Total Revenue

- Sum the revenues from all accepted quotes to determine the total revenue generated.

Iterate and Compare

- Repeat steps 2-4 using various adjustment matrices each time could be scaled up/down by 10%.
- Compare resulting revenues to identify optimal rate configurations.

Conclusion and Future Works

In conclusion, this methodology presents a systematic approach to optimize revenue through simulation and logistic regression modeling. By iteratively adjusting factors and assessing acceptance probabilities, we aim to find the most profitable rate configurations.

1. **Utilize Larger Datasets:** Expand the dataset size to enhance estimation accuracy and robustness.
2. **Validate Logistic Regression Results:** Conduct thorough validation to ensure the reliability and generalizability of the logistic regression model's predictions.
3. **Test on Additional Markets:** Extend testing of the logistic regression model to diverse markets to assess its effectiveness across different regions and scenarios.
4. **Automate Adjustment Factor Refinement:** Develop a systematic approach to automatically refine adjustment factors, streamlining the optimization process and improving efficiency.