

Predicting Flight Delay for August 2022

A penalized regression approach



Xue Li, Ya Wang, Ilias Willems, Jack Xia & Kylan Young

Faculty of Science

Apr 28, 2022

Quick Preprocessing Overview



STEP 01

Aggregate delay
and adjust features



STEP 02

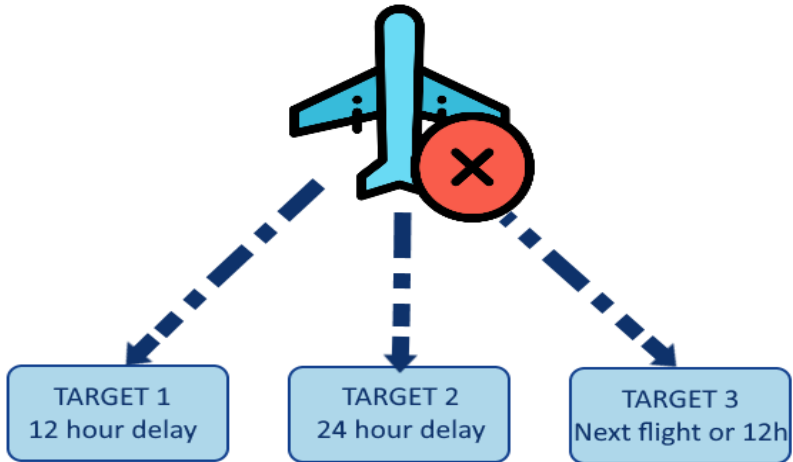
Convert dates
and times



STEP 03

Sort airports into
classes based on
mean delay

What about cancelled flights?



Feature adjustment

Dropped features

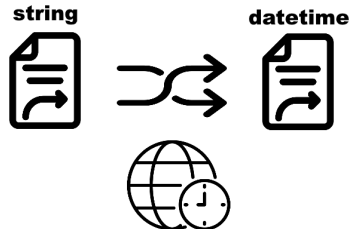
- Delays/Cancelled
- Arrival/Departure times, Taxi/Wheel deployment, Elapsed/Air time
- Flight and Tail number, Cancellation reason

Reason

- Replacement
- Would not be available before departure of future flights
- Irrelevant to the predictions

Date and time conversion

- Workable format
- Adjust to timezone
- Day of the week



Airport classes

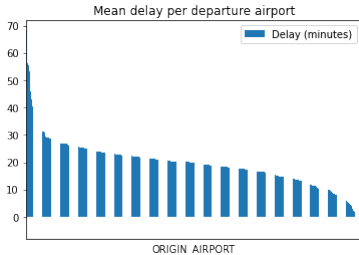


Figure: Mean delay per departure airport

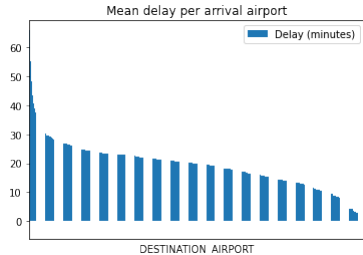
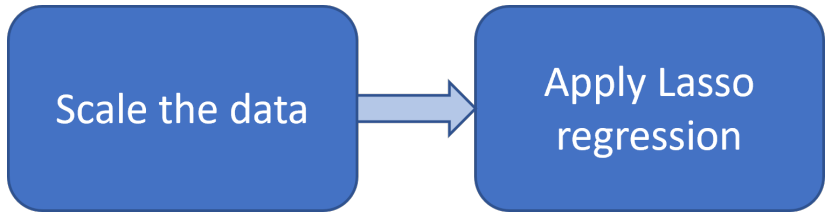


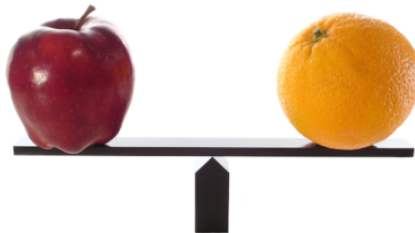
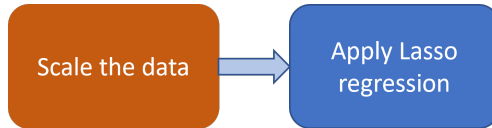
Figure: Mean delay per arrival airport

Top and bottom 5% of airports → High and Low class resp.
Middle 90% and missing obs → Moderate class

The model

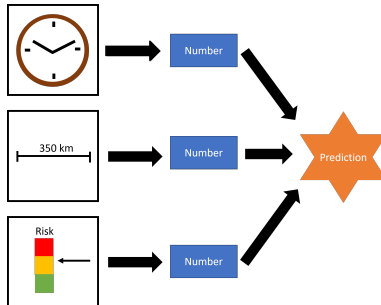
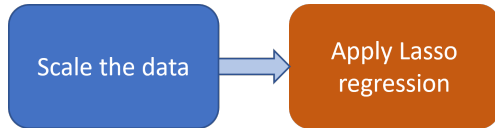


The model



apples-and-oranges. *Locatify*. <https://locatify.com/apples-to-oranges-4/>.

The model



Performance

Performance measures
(on test set)

- $R^2 \approx 0.44$
- $\text{MSE} \approx 6974$

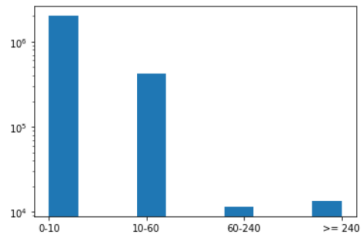


Figure: The distribution of the errors

Prediction

Predictions on delay time (min)

- 0-10 (1.46%)
- 10-30 (65.46%)
- 30-60 (33.03%)
- 60+ (0.05%)

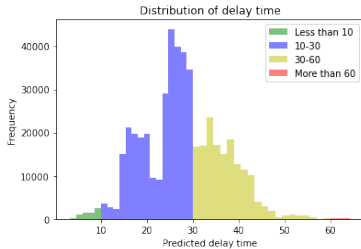


Figure: The distribution of delay time

Prediction

Predictions by airline

- HA: highest (29.6 min)
- ...
- B6: lowest (26.6 min)

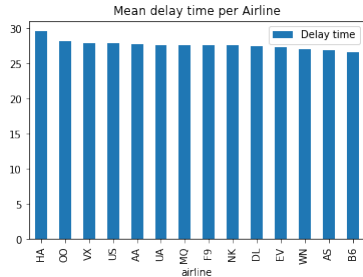
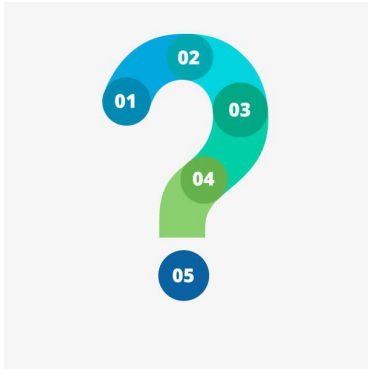


Figure: Mean delay time per airline

Notes and suggestions for improvement



Notes and suggestions for improvement

1. Cancellation reasons?
2. Peak season or off season?
3. Public Holidays?
4. Subjectivity of classes?
5. Group variables?

Conclusion



Takeaways

- Usefulness of this model
- Why did we pick this model?
- Limitations of Model
- Important Assumptions/ Considerations

Next Steps



- Practical Use Cases
- Additional Data supplementation
- Suggested next steps