PREDICTING FLIGHT DELAYS

(A PROOF-OF-CONCEPT)

SOLOMON HENG (*DSI 12*)

Contents:

1. Overview

"Delays are costly for airlines and their passengers. A 2010 study commissioned by the Federal Aviation Administration estimated that flight delays cost the airline industry $8 billion a year, much of it due to increased spending on crews, fuel and maintenance. Delays cost passengers even more — nearly $17 billion."

- https://mashable.com/2014/12/10/cost-of-delayed-flights

1. Overview

MI235 was scheduled to land in Changi Airport at 1235UTC, and planned to park at bay F52, some of its passengers have a connecting flight for SIA21 to the US. However due to a delay at the origin airport, MI235 will be landing at 1310UTC.

SIA21 parked at bay A11 is scheduled to depart at time 1330UTC. The timing was planned with consideration of the connecting passengers from MI235. Because of the delay of MI235, SIA21 is now delayed slightly and will be departing at 1400UTC.

SIA656 is scheduled to land in Changi Airport at time 1330UTC is scheduled to park at bay A11, the bay SIA21 is still occupying and will continue to occupy due to the delayed departure. SIA656 touches down on the runway on time at 1330UTC, and by the time apron control realizes the conflict in the parking bay, the damage is already done. Apron phones ATC to request that they hold the aircraft on the taxiway while they figure out another bay and get the ground handling agents to rush there for the newly arrived aircraft.

(scenario is fictional)

2. Problem Statement

Can we accurately predict flight delays after it has began pushing back from its origin aerodrome?

3. <u>Dataset</u>

1. Kaggle/US DOT Bureau of Transportation Statistics

<u>The dataset consists of 3 files:</u>

i. airlines.csv - Airline information
ii. airports.csv - Airport information
iii. flights.csv - Domestic flight information in US for year 2015

2. OGIMET

The webpage is the source of our weather data. For the purpose of this project, we used the API to retrieve the METAR & SPECI reports for Atlanta International Airport for the year 2015.

4. Key Feature Engineering Processes

NUM_ARR_AVG_3HOUR

This feature will help us understand the competition which a flight faces for landing slots

CROSSWIND_COMP

This feature will help calculate the various crosswind component a flight experiences on landing

5. Methodologies

In summary, the following are the methodologies which we have tried:

1. **Binary Classification Modelling for <u>All Routes</u>** (*followed by Regression for delayed flights*)
2. **Regression Modelling for <u>All Routes</u>**
3. **Regression Modelling for <u>Single Route</u>**
4. **Multiclass Classification Modelling for <u>Single Route</u>** (Final Methodology Deployed)

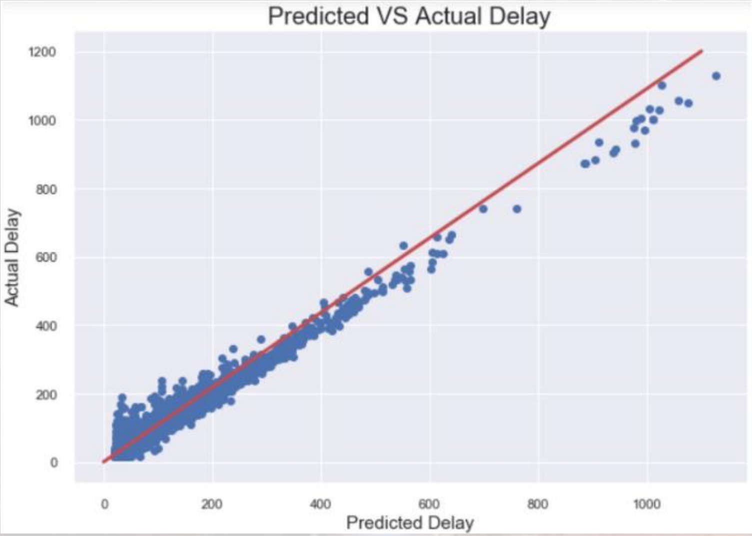
Note: For Single Route modelling, we are only doing it for the top 5 routes with most delays as a proof-of-concept.

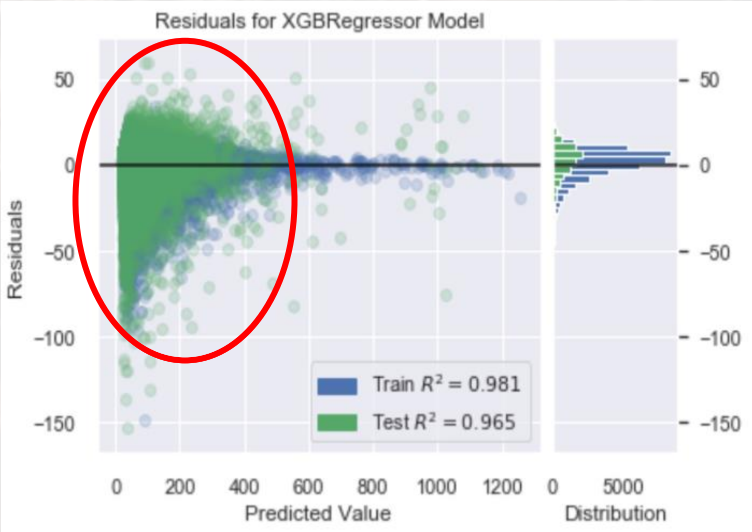
1. Binary Classification Modelling for All Routes

Metric	Score
Sensitivity	77.06%
Specificity	98.99%
AUC-ROC	0.8802
Overall Accuracy	95.82%

2. Regression Modelling for All Routes

Metric	Score
RMSE	15.2 mins
R-Squared	0.9646

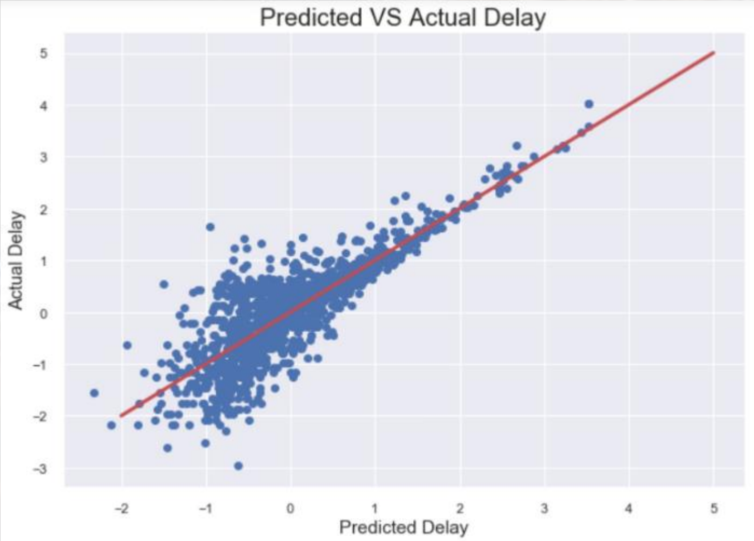


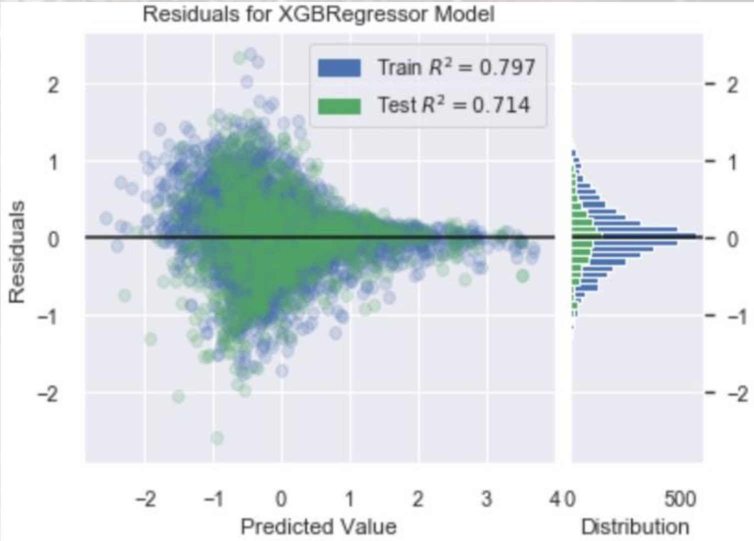


3. Regression Modelling for <u>Single Route</u>

Model & Metric	ORD	LGA	PHL	DFW	MCO
RandomForest RMSE	7.80 mins	7.62 mins	8.38 mins	6.16 mins	2.89 mins
RandomForest R-Squared	0.69622	0.58489	0.54601	0.62830	0.63326
XGBoost RMSE	7.68 mins	7.46 mins	7.99 mins	5.42 mins	2.09 mins
XGBoost R-Squared	0.71376	0.63578	0.55465	0.64771	0.65851
Neural Network RMSE	8.11 mins	9.05 mins	9.09 mins	6.75 mins	2.89 mins

Residual Plots for only ORD





Heteroscedacity

When the ***delays are low****, there* ***are too many possible factors*** *affecting the amount of delays (e.g pilot mistakes, ATC mistakes, another aircraft on the taxiway causing delays). As the* ***delays increase****, the* ***possibilities/noise are "even-ed" out by more stable concrete factors****.*

Hence the decrease of variance in prediction as time proceeds on. I believe this heteroscedastic nature is nearly impossible to handle without decreasing the RMSE of the model (which would likely render the model useless practically).

To work-around this issue:

i. We can perhaps follow a guideline of **only trusting the model when the predictions are above the 1 SD mark** (or when predicted delays are >37mins (*for ORD only, other routes have different threshold for 1 SD*)).

ii. **OR** we can **build a multiclass classification model** to help bin the delays into various periods (***which we will do in our final model***).

6. Final Model

Target Classes & Suggested Purpose:

<15 minutes	***No Delay Category***
15 minutes to 1 hour	***Decide if a reshuffling of ground resource deployment is needed***
1 hour to 3 hours	***Decide on the necessary actions to take to mitigate the impact of the delays*** (e.g. rescheduling transit passengers to another flight to prevent delaying the departure of the connecting flight, etc)
>3 hours	***Decide on how to do "damage control"***

Evaluation Metric for ORD:

	precision	recall	f1-score
<15mins	0.94	0.96	0.95
15mins to 1hr	0.72	0.59	0.65
1 to 3hrs	0.81	0.81	0.81
>3hrs	0.84	0.91	0.87
Overall Accuracy			0.90

Evaluation Metric for DFW:

	precision	recall	f1-score
<15mins	0.96	0.97	0.96
15mins to 1hr	0.68	0.65	0.66
1 to 3hrs	0.80	0.74	0.77
>3hrs	0.81	1.00	0.89
Overall Accuracy			0.92

Evaluation Metric for LGA:

	precision	recall	f1-score
<15mins	0.95	0.99	0.97
15mins to 1hr	0.80	0.59	0.58
1 to 3hrs	0.80	0.75	0.78
>3hrs	0.74	0.92	0.82
Overall Accuracy			0.92

Evaluation Metric for MCO:

	precision	recall	f1-score
<15mins	0.97	0.99	0.98
15mins to 1hr	0.84	0.70	0.76
1 to 3hrs	0.85	0.76	0.80
>3hrs	1.00	0.92	0.96
Overall Accuracy			0.95

Evaluation Metric for PHL:

	precision	recall	f1-score
<15mins	0.93	0.98	0.96
15mins to 1hr	0.75	0.55	0.63
1 to 3hrs	0.83	0.77	0.80
>3hrs	0.87	0.93	0.90
Overall Accuracy			0.91

Best performing model (for the 5 routes):

XGBoost

7. Conclusion & Further Research

Limitations:

The entire model is trained only using 2015 USA domestic flight data. And as such our findings are only limited to the data which we have.

Other potential considerations which we can include in the future are:

i. Including all international flights
ii. Incorporating enroute weather data

Further Research:

A. ***When the delays are low, there are too many possible factors affecting the amount of delays.*** The seemingly heteroscedastic nature of delays has significantly increased the complexity in trying to predict it. Perhaps further studies should be done on it to help us better predict delays in the future.

B. ***The impact of weather on delays seem to diminish with greater distance between the 2 airports.*** The hypothesis makes logical sense as the further the distance, the more "buffer" the flight has to make up for time and as such would be less affected by approach weather. Further studies could be done on this to help us understand the impact of weather on delays more.

8. Sample Deployment