Predicting Flight Delays Using Neural Networks

Garrett Tate, April 2021

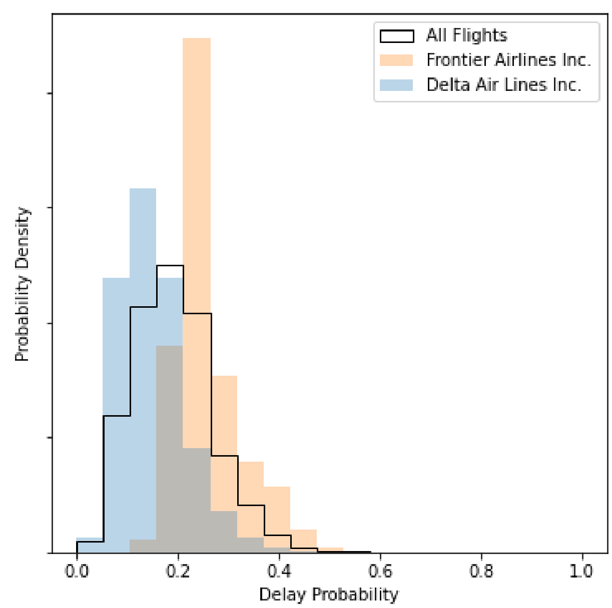
Delayed flights present a significant challenge to travelers and airlines alike, with 19% of US domestic flights delayed in 2019[1] at an estimated total cost of $33 billion that year[2]. Costs to airlines range from crew overages to buffers reducing aircraft utilization, and costs to passengers include actual missed connections as well as travelling earlier than necessary in anticipation of delays. Meanwhile, approaches to plan for flight delays remain rudimentary and inefficient, with airlines applying blanket minimum connection times and fliers often relying on anecdotal experience for delay probabilities. There exists significant opportunity to improve the efficiency of airline travel by estimating the probabilities of delays for individual flights, thereby allowing airlines to schedule more efficiently, third-party bookers to better serve customers with more reliable itineraries, and consumers to plan critical trips more confidently.

Here I have constructed a neural network modeling procedure that predicts the probability of flight delays 1-2 months in advance using basic information known at the time of booking. An in-depth summary and the full Python code for modeling and evaluation can be found at github.com/Tate-G/portfolio. I train and test these models using data from the Reporting Carrier On-Time Performance database from the US Bureau of Transportation Statistics[3].

My neural network modeling procedure provides significant improvement for predicting individual flight delays compared to a baseline of assuming a constant delay probability. Model ROC area under the curve (AUC) ranged from 0.62 – 0.69 when tested against each month in 2019 compared to a baseline AUC of 0.50. In 2020 during the COVID-19 pandemic this modeling encountered more muted success predicting delayed flights, but model performance was at least as good and usually better than the baseline for all but one month in 2020 despite radical changes in air travel.

My modeling allows for significant flexibility if deployed, favoring either precision or recall as best suits the application. For instance, airlines might favor a model that maximizes precision by identifying specific flights that are very likely to be delayed, while customers may prefer models maximizing recall that identify as many reasonably likely delays as possible. While testing against 2019 data, models favoring precision correctly identified an average of 338 delayed flights per month with average precision of 52%, while models favoring recall averaged 62% recall and correctly identified an average of 61,819 delayed flights a month.

I also identify the most important factors for predicting flight delays using a custom-built permutation feature importance[4] procedure. Significant factors include departure/arrival times, month, key airports, and certain airlines, for example the distinct differences in delay probability for Frontier and Delta seen here.



Opportunities abound for impactful real-world deployment of flight delay models such as this. Airlines can reschedule flights at high risk of delay, or booking services can incorporate delay probabilities into connection times. As we exit the pandemic, now is the time to make schedules more efficient, capture customer loyalty, and save billions of dollars for airlines and passengers alike.

---

1 United States Bureau of Transportation Statistics, see data here

2 United States Federal Aviation Administration, Cost of Delay Estimates 2019, available here

3 Data download available here

4 Based on Molnar, 2021, Interpretable Machine Learning, available here