

# Predicting Flight Departure Delay - Team: RooKeys

## Executive Summary

### Overview

A delayed flight causes major financial losses to airline companies, airports, and travelers. For instance, expenses incurred by a traveler for accommodation and food due to flight delays and/or missed connections together with time lost and being away from home. This can result in a decrease in air travel demand from existing and potential customers for an airline because travelers bank on efficient service quality and performance. Impact of delays can translate into productivity slowdown, indirectly affecting the economy and stunting GDP.

### Problem

Keeping this in mind, our aim is to build a classifier model to answer the following question:

- Given a 4-hour period horizon, will a flight be delayed by more than 15 minutes?
- Our target audience is airline companies, as according to the Federal Aviation Administration, the total flight delays cost is \$22 billion yearly.

### Recommended Solution

We aim to put emphasis on “capturing more delays” since to classify a delay incorrectly is costlier for airlines.

- We built a model with an ability of 80% and a reliability of 40% to predict a delay.
- Complexity does not always guarantee performance; our top models were logistic regression and random forest (very slightly better).

This will help airline companies to efficiently manage their operations for expected delays and improve the quality of service to customers.

### Highlight

We performed a rolling average analysis to capture the spatial and temporal correlation. In particular, we found that the delay happens in the clusters across spatial and temporal dimensions (as shown in the figure).

### Conclusion

We embedded the historical information into new probabilistic features, for instance, delay for a given carrier exhibits seasonality and drift. Capturing this information was challenging and important to account for business specific developments associated with the airline industry. This increased the performance of our baseline model significantly providing a simplest solution to the stated problem.

- We found feature engineering to be a critical step for model performance.

