Project McNulty Proposal
Predicting An-Time Flight Status

Dana Lindquist
October 17, 2018

Airline flight on-time status is a big deal for both passengers and airlines. Airlines are penalized in scores that measure on-time rates. Passengers want to get to their destinations as planned.

The goal of this project is to predict if a flight will be on-time or late. The standard definition of on-time is departing or arriving within 15 minutes of the scheduled departure or arrival time.

To simplify the project, I will focus on domestic flights arriving at and departing from SeaTac International Airport in Seattle, WA. Weather is a clear reason for delays, and not just the weather in Seattle. To account for weather in other parts of the US I will include weather in other US cities as features.

The data I will collect is in the following table. Data for training will be pulled from June 2017 to July 2018, the most recent reporting data for the Bureau of Transportation Statistics.

Variable	Type	Description	Use in Model
Flight Number	int		N
Airline Carrier	int	Will be modeled with one-hot encoding (one column per airline)	Y
Date	datetime		Y
Arrival or Departure	boolean	1 for arrival, 0 for departure	Y
Planned Arrival Time	datetime	Arrival at Seattle for inbound flights Arrival at destination city for outbound flights	Y
Actual Arrival Time	datetime		Y
Flight Distance	int	Distance in miles from Seattle to city of flight destination or departure	Y
Weather in Seattle - Average Daily Temperature Average Wind Speed % Precipitation	 Int Int int	Representation of weather in Seattle. These items are available as a forecast which would be an input for the prediction. More information than remote cities.	Y
Weather in Chicago - Average Daily Temperature Average Wind Speed	 Int int	Representation of weather in remote city. These items are available as a forecast which would be an input for the prediction.	Y
Weather in NYC - Average Daily Temperature	 Int		Y

Average Wind Speed	int		
Weather in LA - Average Daily Temperature Average Wind Speed	 Int int		Y
Weather in Houston - Average Daily Temperature Average Wind Speed	 Int int		Y
Weather in Atlanta - Average Daily Temperature Average Wind Speed	 Int int		Y
On-Time or Late			**TARGET**

Issues to consider:

- While much weather data is available it will be interesting to see what is actually important. I foresee some exploration in this area.
- There are approximately 30,000 data points per month. I will have to consider the dataset size. I might have to focus on just outbound or inbound flights

Sources of data:

Airline On-Time statistics - Bureau of Transportation Statistics
https://www.transtats.bts.gov/Tables.asp?DB_ID=120&DB_Name=Airline%20On-Time%20Performance%20Data&DB_Short_Name=On-Time

Weather data for cities - NOAA
https://www.ncdc.noaa.gov/cdo-web/datatools/lcd