Miguel Montano
Capstone Project 1 Proposal

Predicting the Occurrence of Flight Delays

Problem: Predicting the instance of an arrival delay based on a flight's departure delay

Client Interest: According to a report from the leading Airline Industry advocacy group, Airlines For America, in 2016 the average cost of aircraft block (taxi plus airborne) time for U.S. passenger airlines was $62.55 per minute. Further analysis has revealed that delayed aircraft are estimated to have cost airlines several billion dollars in additional expenses annually. Delays subsequently result in the need for additional gates and ground personnel, and impose costs on airline customers (including shippers) in the form of lost productivity, and additional wages. In 2010, FAA/Nextor completed a comprehensive study on the costs and impacts of flight delays in the U.S. and estimated the annual costs of delays in 2007 to be **$**31 billion ("Per-Minute Cost of Delays to US Airlines", airlines.org). A tremendous amount of capital is spent annually on mitigating the effects of delay, but with an ability to predict the likelihood of an arrival delay based on a departure delay, further steps can be taken to anticipate necessary remedies to prevent auxiliary costs.

Data Source and Acquisition: Data has been acquired from an online posting on Kaggle containing US domestic flight delays in the month of January,August, November and December of 2016. Data is free and has no contingencies upon its usage, and contains extensive descriptive data on all domestic flights within the time range, including their arrival/departure locations, time intervals spent on various stages of the arrival/departure process, and arrival/departure deviations from their scheduled times, which provides the incidence of delay.

Approach: Flights will be filtered into two categories, delayed or not delayed, based on their deviation from their planned times, and then the departure and arrival components will be used to build a model based on their delay status. A logistic regression and other algorithms to be determined will then be used to predict of the delay status of future flights.

Deliverables: A model capable of providing predictions as to a whether or not an arriving flight will be delayed based on whether or not it departed with a delay.