

# CAS CS 505 Final Project Proposal

(Topic: Prediction of Flight's Departure Performance)

**Tian Zhang, Lu Min**

## 1. Dataset

### 1.1 Dataset Introduction

For getting valuable prediction result, we decide to collect dataset from the official website of Bureau of Transportation Statistics (BTS) which offer us publicly-available data on flight arrivals and departures for major U.S. airports. The downloaded dataset contains 29 variables that describe each flight in terms of departure/arrival date and time, carrier, taxi time, time spent in the air, as well as departure and arrival delays and their causes. Variables as the following form listed:

TailNum	Year	TaxiOut
ActualElapsedTime	Month	Cancelled
CRSElapsedTime	DayofMonth	CancellationCode
AirTime	DayOfWeek	Diverted
ArrDelay	DepTime	CarrierDelay
DepDelay	CRSDepTime	WeatherDelay
Origin	ArrTime	NASDelay
Dest	CRSArrTime	SecurityDelay
Distance	UniqueCarrier	LateAircraftDelay
TaxiIn	FlightNum	

The dataset is split into individual files, on file per month, and contains around millions of records. Each record represents a unique flight record as a combination of a flight number and aircraft tail number.

### 1.2 Data Processing

Practically, we do not concern all features in our project, so we decide to work with the most relevant variables to build our models, like the data concerning about four categories, and the choosing reasons are described:

- Time variables: the daily, weekly, monthly, seasonal and annual fluctuation in flight

will have directly effect in flight delays.

- Origin and Destination variables: the working load or efficiency of airport will also influence the departure performance of flight.
- Aircraft Tail Number: the technique difference between aircrafts will also cause different departure performance.
- Airline Carrier variables: it is possible that some airlines have higher propensity for departure delays.

And we will use binary classification to label every data record: '1' is used to label the flight delays ('Delay' is defined as that the departure of flight is late for more than 15 minutes), and '0' presents the on-time flight.

## **2. Expected Analyses**

In one word, our project can be concluded as a classification problem, since we will create a model for analyzing flight delays, and then use such model to make classification for new data. Thus, we consider several algorithms including Decision Tree, Naïve Bayes, SVM, and Logistic Regression. And we believe that the prediction of flight delays is satisfied to the Markov assumption due to the reliance of a current flight's departure on the airplane arriving on-time from its previous flight. And finally, we will figure which model is more appropriate for our project.

## **3. Application**

According to our anticipative analysis, we will predict the delay status of airlines from different airports with the same destination and provide a flight with less probability to delay.

There is no actual application associated with our dataset, and we think our analysis is beneficial for business users who know that they will travel on a certain date, but are uncertain as to the probabilities on that date of a flight delay. If passengers can know which flight would be more possible to delay, they would choose the flight with less possibility to delay for saving their time and money.

## **4. Expected results**

First, we expect to predict the delay status of flights from different airport with the same destination and departure date. Second, we would analyze the factors

which attribute to the delay, and figure out in which season or which quarter of a year. Finally, we can obtain a record with the lowest probability to delay with given time period and destination. Each record contains a unique flight record as a combination of a flight number and aircraft tail number.

Since we would use several models to obtain our results, we would also we will make conclusions about the suitability of our models for delay predictions. We would then visualize a certain year with which is the preferred airport from which to depart for a few selected destinations and given departure date.