**Tweeter US Airline Sentiment**

**Problem and its relevance to airline business.**

The most important success factor for airlines is the credibility they offer to their customers. This indicator is expressed in terms of customer satisfaction. There are many factors that can change this satisfaction either positively or negatively. Therefore, it is necessary for an airline to know the factors that affect these sentiments.

**The data.**

Twitter data was scraped from February of 2015 and was firstly asked to classify positive, negative, and neutral tweets, followed by categorizing negative reasons (such as "late flight" or "rude service").

**The model**

First we need to pre-process the data. We need to handle missing values, outliers, noisy and high correlation data. Then,it is important to process the content of Tweets that they express the sentiments of customers. Τhis process is known as natural language processing (NLP).

We promise to apply classification models and neural networks to find out which is the best fit for our problem.

**How the model will be evaluated**

We will evaluate the performance of this method using a Kaggle test data set. The metric we will use is the prediction error of a customer’s sentiment(positive, negative,neutral tweets).

**Anticipated challenges**

It is likely that the "tweets sentiments" (and mainly the negative) of some customers are not related to the attributes given by the airlines and can be purely personal. Ηowever, we will perform model selection and compare the performance of joint model vs. separate models for the two datasets. Regardless of outcome, this will result in useful observations about the airline services that can help them to improve as companies.

**The promise**

Given the increasing competition between airline companies, the understanding of passenger qualities will help increase the credibility of each company.