

US airline Twitter data

Airplane travel is a huge business in the US. People are always talking about their air travel stories to each other being good or bad. With twitter people can and will tag the airline in their response to their travel.

Since Air travel is a service industry Airlines will want to know how they can improve or what they need to do better from their customers. From twitter data the airlines can see what features are causing positive or negative tweets. And from the data decide what to do as a company.

The data was collected from Twitter in February 2015. The contributors were asked to first classify positive, negative, or neutral tweets. The data is on kaggle at <https://www.kaggle.com/crowdflower/twitter-airline-sentiment>.

To solve the problem a model will be built to analyze the tweets text to see if the model can predict a positive, neutral, or negative tweet. Bag of words method will be used with n-gram to see if there are different phrases that will help. There will be three different algorithms used to build the model: Decision Tree, SVC, and MultinomialNB.

The deliverables will be a jupyter notebook with the code, and a paper describing the process and results of the analysis.