Sentiment Analysis of Airline Tweets

Rika Yatchak

March 16, 2019

1   Problem Description

We have a dataset of tweets containing @ mentions of various U.S. domestic airlines such as United, Virgin America, and others. Is a given tweet positive, negative, or neutral in sentiment?

2   Client

The client is a U.S. airline that is interested in improving its public image. They would like to prioritize replying to tweets from unhappy customers in the interest of increasing their satisfaction.

3   Data

We will work with the Twitter US Airline Sentiment dataset, which was originally collected by Crowdflower (now Figure Eight) and uploaded to Kaggle by Figure Eight.

This dataset contains 14,000 individual tweets and their associated metadata (id, user timezone, time posted, etc). It also contains an airline sentiment label (positive, negative, neutral).

4   Approach to the Problem

We will use the traditional Bag-of-Words model, the TF-IDF model, and Doc2Vec. After representing indivudal tweets using the schema in each of these models, we will apply logisitic regression and random forest to solve the sentiment classification problem.

5   Deliverables

Code for the project will be available on github, and a presentation with summary results will be provided.