

Twitter US Airlines sentiment

1. Project idea

In our study we want to analyse how the customer-company relationship impacts on the revenue of each company as well as on the stock market. We will try to demonstrate empirically whether a negative sentiment analysis on Twitter for a given airline negatively affects the stock price and vice versa.

2. Database

Twitter data was scraped from February of 2015 and contributors were asked to first classify positive, negative, and neutral tweets, followed by categorizing negative reasons (such as "late flight" or "rude service"). Nevertheless, we also consider to develop our sentiment classifier since we think will represent a more reliable approach.

We will use the following variables:

2.1) Kaggle Database:

- Tweet_id: The primary key
- Airline_sentiment: Positive or negative (boolean)
- Airline_sentiment_confidence: Logistic regression (I guess.. since values form 0 to 1)
- Negativereason: Negative topic (i.e.: delay, bad flight, customer service...)
- Airline: The Airline's name (American Airlines, United Airlines, US Airways Southwest, Delta and Virgin America)
- Name: The user's account name (Some users tweeted more than once)
- Retweet_count: Number of the retweets
- Text: The tweet
- Tweet_coord: Coordinations of the tweet
- Tweet_created: When the tweet was created (timestamp)
- Tweet_location: Location of the tweet (i.e.: Los Angeles, San Francisco...)

2.2) Yahoo Database:

- Date: From 01/02/2015 to 31/12/2015
- Open: Open price.
- High: Highest price during open hours.
- Low: Lowest price during open hours.
- Close: Close price.
- Adj Close: Adjusted close price.
- Volume: Daily volume.

Notes:

1. Since not all companies are traded, we consider the possibility to add more financial information, for instance from accounting statements: assets to liabilities, debt cost, profit margin...
2. Another limitation is that the tweets were gathered from 17/02/2015 to 24/02/2015 (8 days). However the total observations are 15.000 tweets. We are considering to include other review sources.

Some exploratory ideas on this:

- When tweets are negative, what topics do travelers tend to discuss?
- When tweets are positive, what are travelers happy with?

- Outside of sentiment, what other systematic variation exists in the tweets about different airlines?
- What types of tweets on airlines tend to be retweeted?

Ahmad proposition - > **Clustering negative tweets on topics** (topic classification)

Databases links:

Kaggle: <https://www.kaggle.com/crowdflower/twitter-airline-sentiment>

American Airlines Group Inc. (AAL):

<https://finance.yahoo.com/quote/AAL/history?period1=1422745200&period2=1451516400&interval=1d&filter=history&frequency=1d>

United Airlines Holdings, Inc. (UAL):

<https://finance.yahoo.com/quote/UAL/history?p=UAL&.tsrc=fin-srch>

Southwest Airlines Co. (LUV):

<https://finance.yahoo.com/quote/LUV/history?period1=1420066800&period2=1451516400&interval=1d&filter=history&frequency=1d>