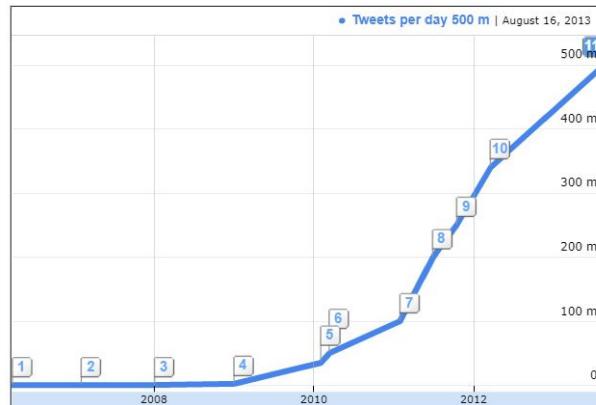


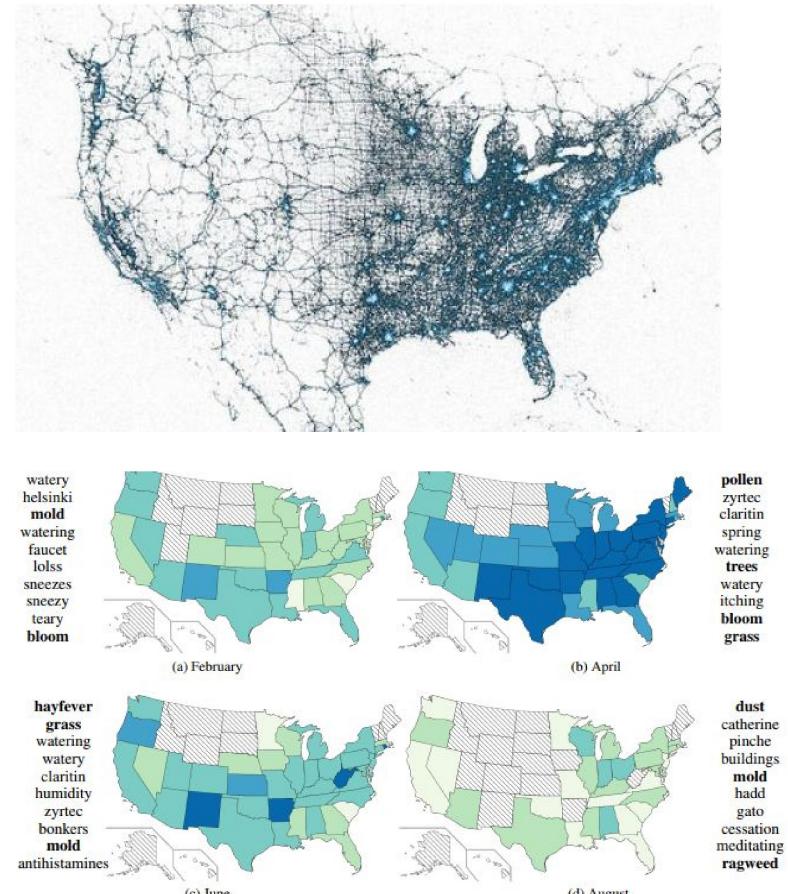
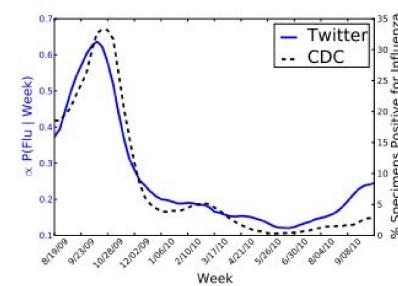
The Value of Twitter

- Social network with 280 character messages called “tweets”
- 310 million monthly users
- 500 million tweets per day
- Delivers a real-time picture of what’s happening around the world at any minute
- Gain overview of the wider public opinion behind certain topics



What kind of data does Twitter have?

- People hand over their data
 - Location
 - Status
 - Opinions
 - Ideologies
- Business, Political, Public opinion mining
- Trends
- Message spread (retweet amplification)
- What other kinds of data?
 - Health: “I caught the flu ☺”
 - Weather i.e. “social weather mapping”



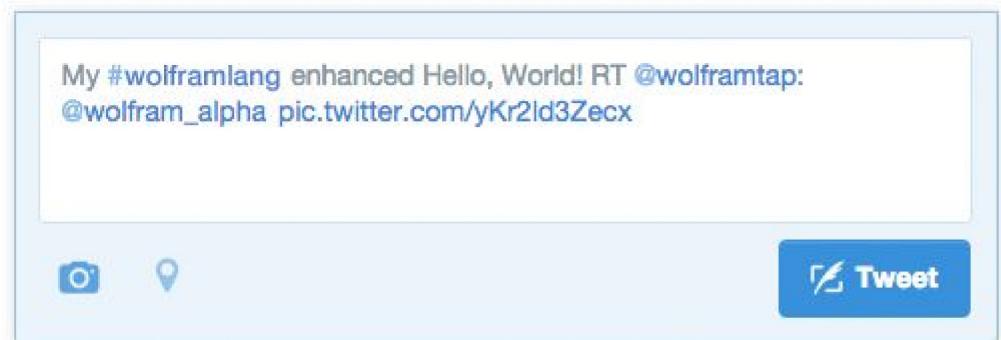
Deriving Meaning from Tweets

- Natural language processing enables computers to derive meaning from human language
 - Sentiment Analysis collects and classifies opinions from text
 - Positive, negative, or neutral
 - Can be used by businesses to gauge attitudes towards products
 - 19% of tweets mention brands, 20% of those tweets showed sentiment towards the brand
 - Can also predict shifts in markets



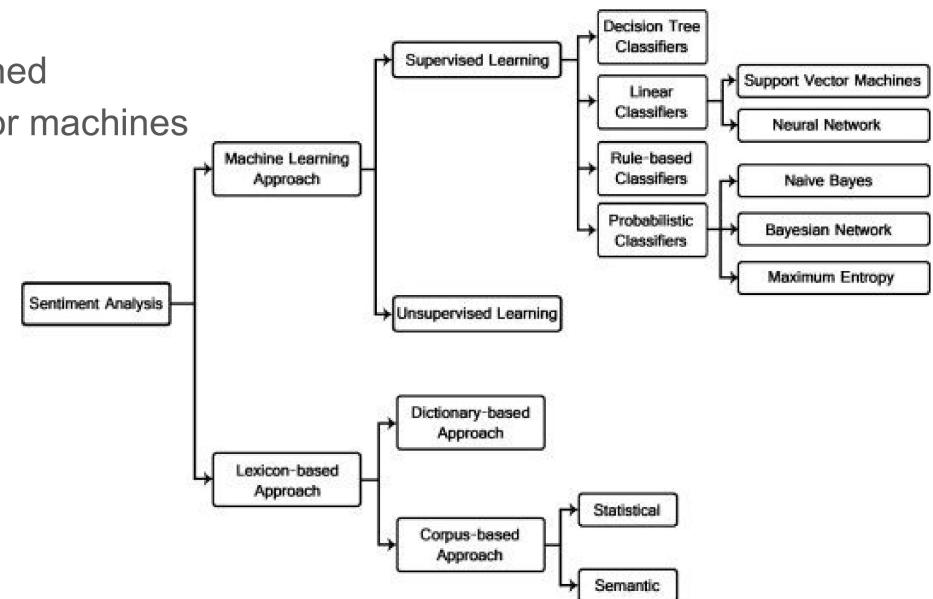
Challenges of Sentiment Analysis of Twitter

- Length
 - About a few sentences or less
- Informal language
 - Slang
 - Misspelled words
 - Does not conform to grammatical rules
 - Symbols, links, emoticons, hashtags
- Magnitude of data set
 - 6,000 tweets sent every second
 - Utilize the Twitter API



Approaches

- Lexicon
 - Words and emoticons given polarity scores. Sentiment of tweet based upon average score of words/phrases
- Machine Learning
 - Many features selected and model is trained
 - Naive bayes, decision trees,support vector machines
- Combination Approach
 - Lexical analysis used as input feature to machine learning approach
 - Most effective

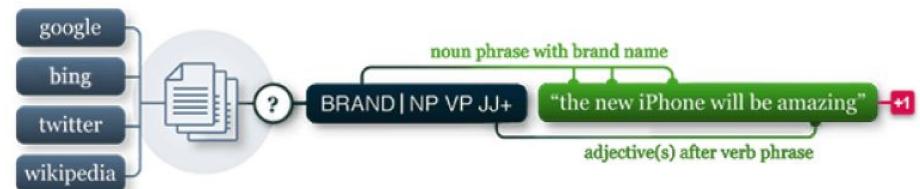


Tools

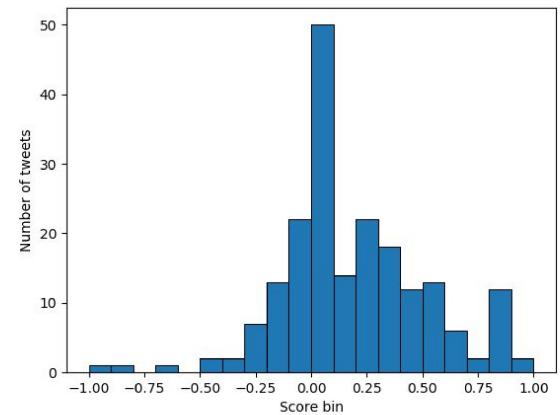
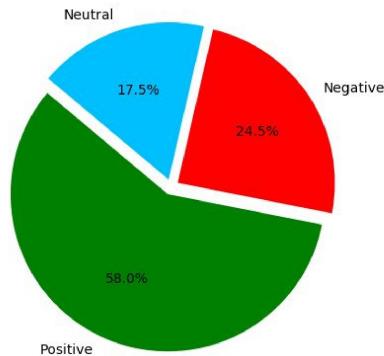
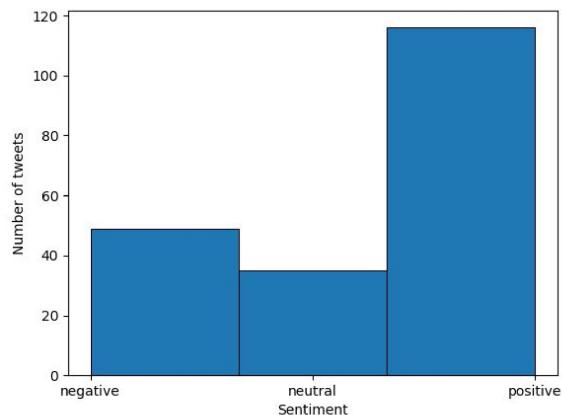
- Tweepy
 - Tweets, followers, querying, live streaming, etc ...
- NLTK corpora
 - Training data
- Matplotlib, NumPy
- TextBlob
 - Sentiment analysis
 - POS, tokenization, frequencies, word inflection etc
 - Pattern Analyzer, Naive Bayes, Maximum Entropy etc.

Methodology

- Authentication
- Fetching
- Pre-Processing
- Sentiment Analysis
- Output



@realDonaldTrump, Top 200



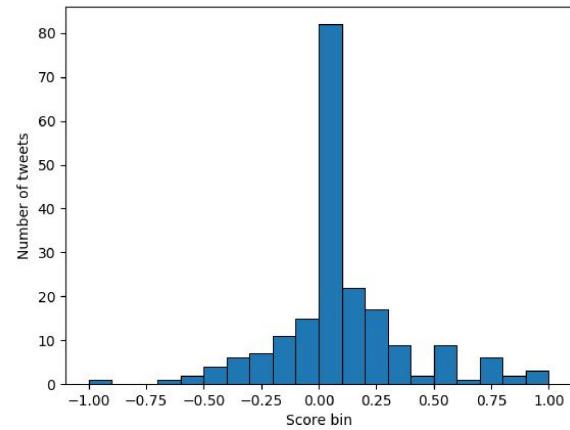
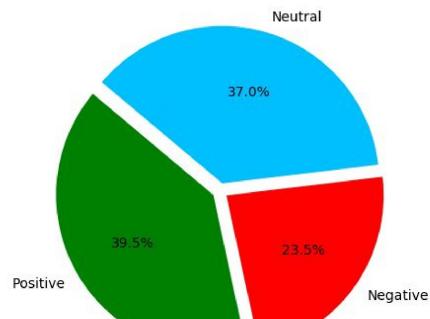
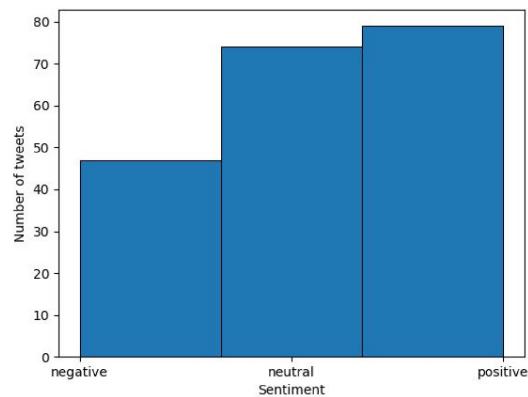
Positive tweets:

- (1) We had a wonderful visit to Vietnam, thank you President Tran Dai Quang! Heading to the #ASEANSummit 50th Anniv Gala in the Philippines now.
- (2) Playing golf with Prime Minister Abe and Hideki Matsuyama, two wonderful people!
- (3) Just returned from Asia after 12 very successful days. Great to be home!
- (4) It was my great honor to have lunch with our INCREDIBLE U.S. and ROK troops at Camp Humphreys, in South Korea.
- (5) Thank you to President Moon of South Korea for the beautiful welcoming ceremony. It will always be remembered.

Negative tweets:

- (1) Horrible and cowardly terrorist attack on innocent and defenseless worshipers in Egypt. The world cannot tolerate terrorism, we must defeat them militarily and discredit the extremist ideology that forms the basis of their existence!
- (2) Will be calling the President of Egypt in a short while to discuss the tragic terrorist attack, with so much loss of life. We have to get TOUGHER AND SMARTER than ever before, and we will. Need the WALL, need the BAN! God bless the people of Egypt.
- (3) The good news is that their ratings are terrible, nobody cares!
- (4) We should have a contest as to which of the Networks, plus CNN and not including Fox, is the most dishonest, corrupt and/or distorted in its political coverage of your favorite President (me). They are all bad. Winner to receive the FAKE NEWS TROPHY!
- (5) The last thing we need in Alabama and the U.S. Senate is a Schumer/Pelosi puppet who is WEAK on Crime, WEAK on the Border, Bad for our Military and our great Vets, Bad for our 2nd Amendment, AND WANTS TO RAISES TAXES TO THE SKY. Jones would be a disaster!

@TheOnion, Top 200

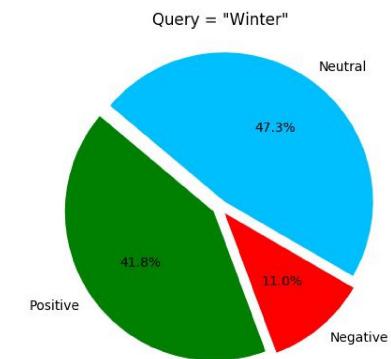
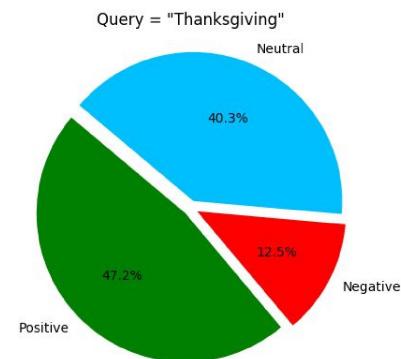
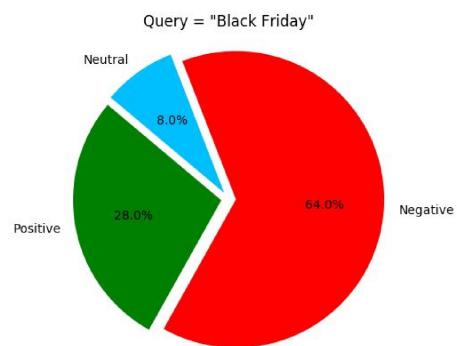
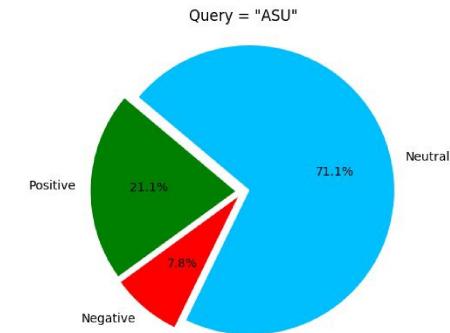
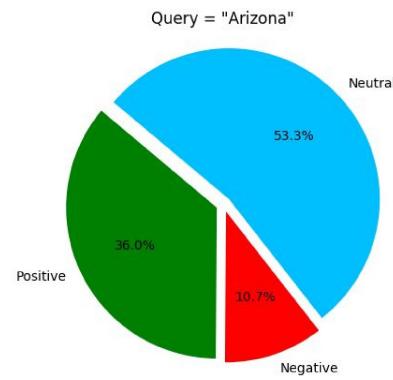
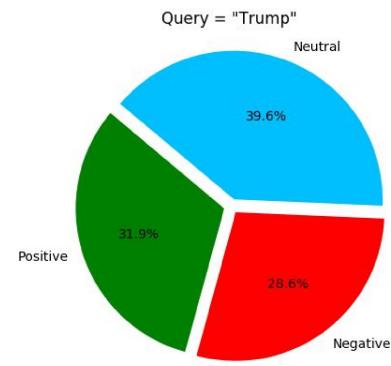


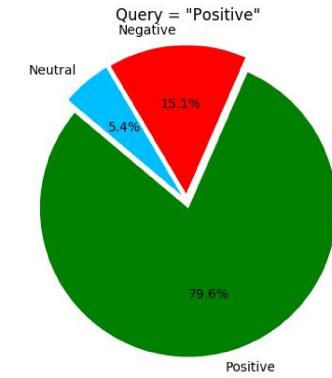
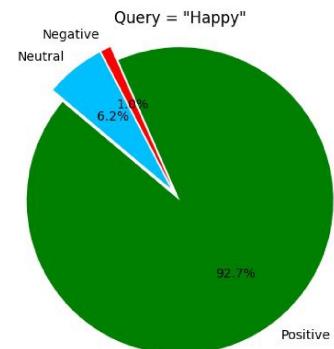
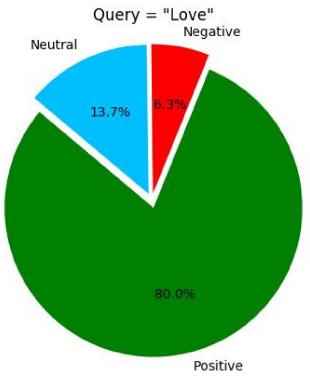
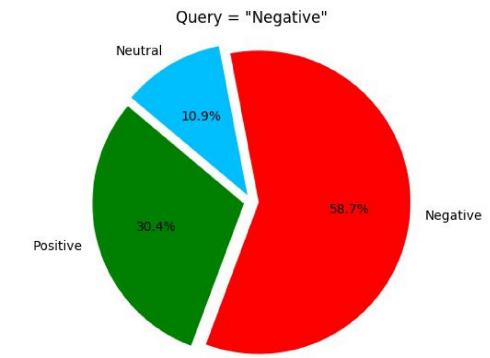
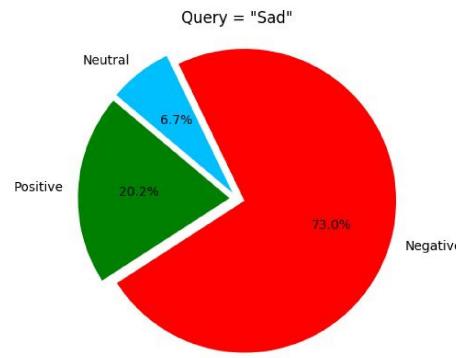
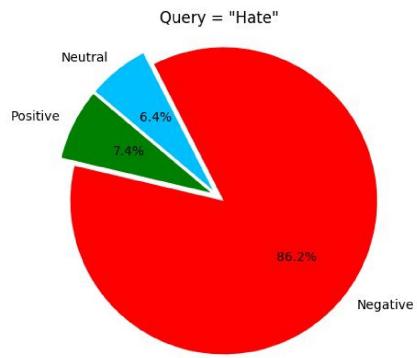
Positive tweets:

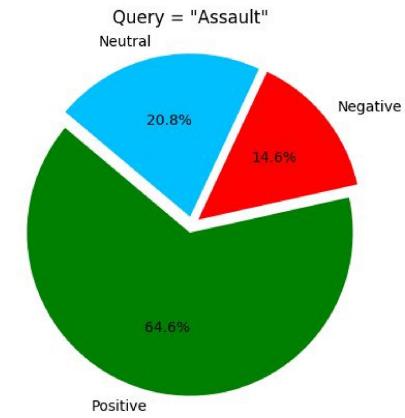
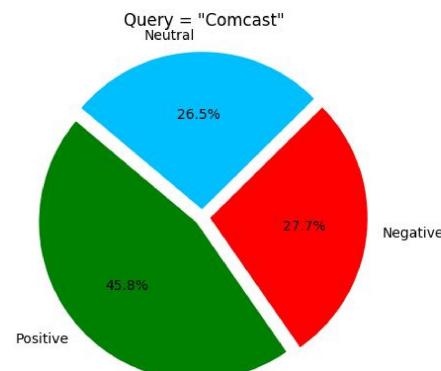
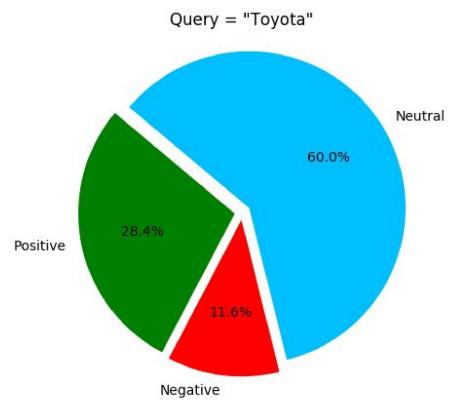
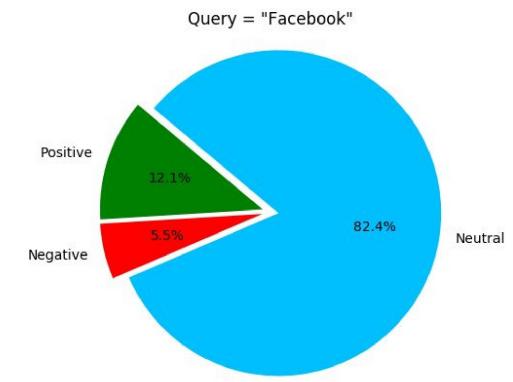
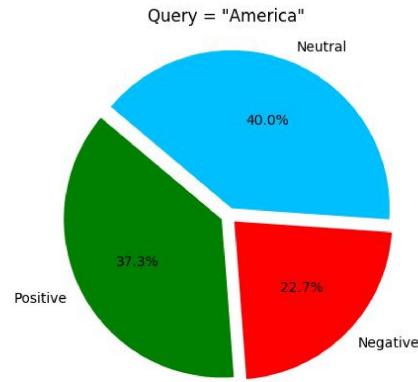
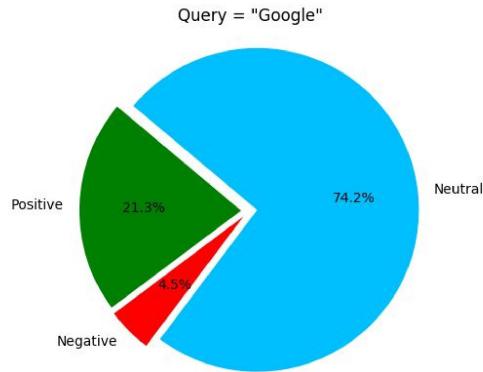
- (1) Attempt To Recreate Incredible Night Out From Youth Works Perfectly
- (2) Hank's Having A Great Time In The Studio Alone On Thanksgiving
- (3) Showerin' Real Good Continues To Top Bridal Style Trends Of 2017
- (4) Unclear Whether Grandpa Having Good Time
- (5) RT @ClickHole: Genius Marketing: Poland Spring Is Putting Apple Juice In Its Bottles Because It Tastes Better

Negative tweets:

- (1) Papa John's Comes Under Fire For Cruel Treatment Of The Bulbous, Deformed Creatures That Lactate Pizza Sauce
- (2) RT @ClickHole: Get F***** Up On Gin And Build A Gazebo For Your Neighbor
- (3) Macy's Parade Float Covered In Tickets After Parking On 5th Avenue Over Holiday Weekend
- (4) Woman Drawn To Shampoo With Most Gruesome Description Of Hair
- (5) Dying Woman Sorry She Won't Get To See 37-Year-Old Son Grow Up







Improvements

- Improve preprocessing
 - remove stop words
 - keep emoticons
 - split hashtags into separate words
- Use comprehensive training data
 - Test different training sets
- Improve classifier
 - Compositionality
 - Measure subjectivity
 - Categories outside of pos/neg/neu
- Visualization
 - Word bubbles, term frequencies
 - Sentiment maps

References

- http://www.cs.jhu.edu/~mpaul/files/2011.icwsm.twitter_health.pdf
- https://blog.twitter.com/marketing/en_us/a/2014/study-exposure-to-brand-tweets-drives-consumers-to-take-action-both-on-and-off-twitter.html
- <http://www.ijcaonline.org/research/volume125/number3/dandrea-2015-ijca-905866.pdf>
- <https://www.clips.uantwerpen.be/pages/pattern-en#sentiment>