**Twitter Sentiments Analysis using Python Model**

* Sentiment analysis (or opinion mining) : uses natural language processing and text analysis to determine author’s attitude towards a particular topic in a piece of text

* Basic tasks:
  + Determine polarity (positive/negative/neutral)
  + Determine subjectivity/objectivity

* More advanced tasks include examining emotional states such as anger, sadness, happiness

# Opinion Mining in Different Industries

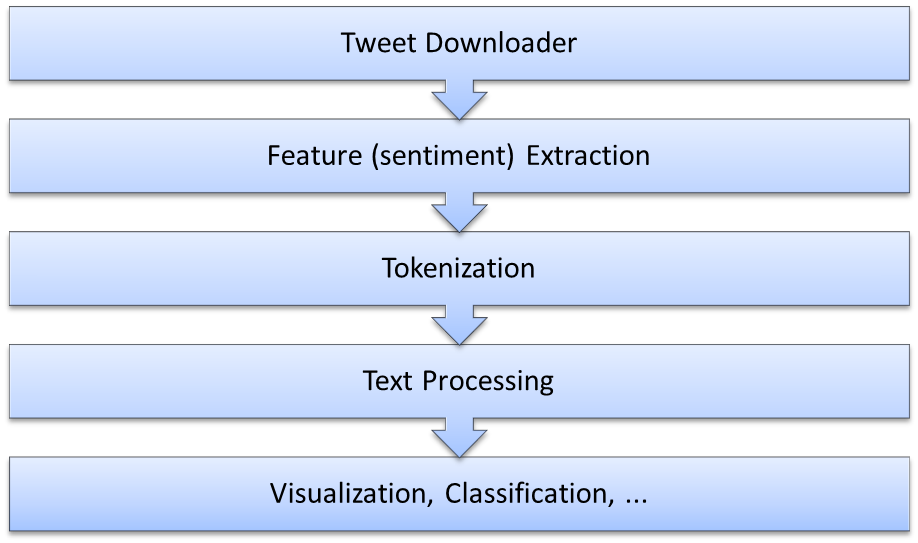
* American Red Cross uses the SaaS tools of Radian6 to monitor social media comments made by its volunteers and donors (in addition to survey and inperson focus groups)

* The Wall Street Journal’s Sentiment Tracker tracks Facebook & Twitter users. They share their findings not as scientific public opinion polls, however.

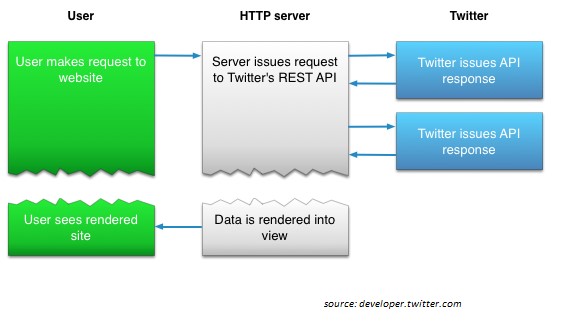
* Dell’s ‘Social Media Listening Command Center’ responds to service related questions and complaints and monitors for consumer trends

* Proctor & Gamble, American Express, DirecTV other corporations with ‘listening centers’

# Approach



**Tweet Downloader (through Twitter API)**



# Feature (sentiment) Extraction

* Classical machine learning methods to do polarity classification heavily dependent on training data

* Other methods use external lexical resources like WordNet, or Sent WordNet that identify polarity of words along with intensity

* Building a Twitter classifier model not the goal in this presentation (we use an existing RapidMiner operator for this)

# Challenges of Sentiment Extraction

* What a pronoun, or a noun phrase refers to. "We watched the movie and went to dinner; it was awful." What does "It" refer to?

* Parsing - What is the subject and object of the sentence, which one does the verb and/or adjective actually refer to?

* Sarcasm - If you don't know the author you have no idea whether 'bad' means bad or good.

* Twitter - acronyms, lack of capitals, poor spelling, poor punctuation, poor grammar

* Detecting more in depth sentiment/emotion (beyond positive/negative): how much hate there is inside an opinion….

# Tokenization/Text Processing

* Tokenize: split the document’s text into a sequence of tokens (words, sentences, etc.)

* Filter tokens: exclude certain common words, English stop words, etc…

* Choose a schema for processing Tweets (TF/IDF, Term Frequency, Term Occurrence)

* Use the created vectors (list of words) along with Tweet sentiments in Tableau

# Next…

* Visualization
  + Format the data
  + Build Tableau dashboards at both Tweet and word levels
* Cluster Analysis
  + Use the vector of words (or select subset of them) to form clusters of tweets such that tweets in cluster are similar to each
  + other and are dissimilar to tweets in other clusters

## Final thoughts

* Enormous amount of unstructured data available

* Automatic Sentiment Analysis challenging by nature; no perfect tool (yet!)

* No expertise required for basic Sentiment Analysis