CALIFORNIA STATE UNIVERSITY FULLERTON

Campus Sentiment Analysis Using Twitter Data

Afshin Karimi
Sunny Moon
Rohit Murarka

Office of Assessment and Institutional Effectiveness
CSU Fullerton

2017 CAIR Conference - Concord, CA

Sentiment Analysis

- 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

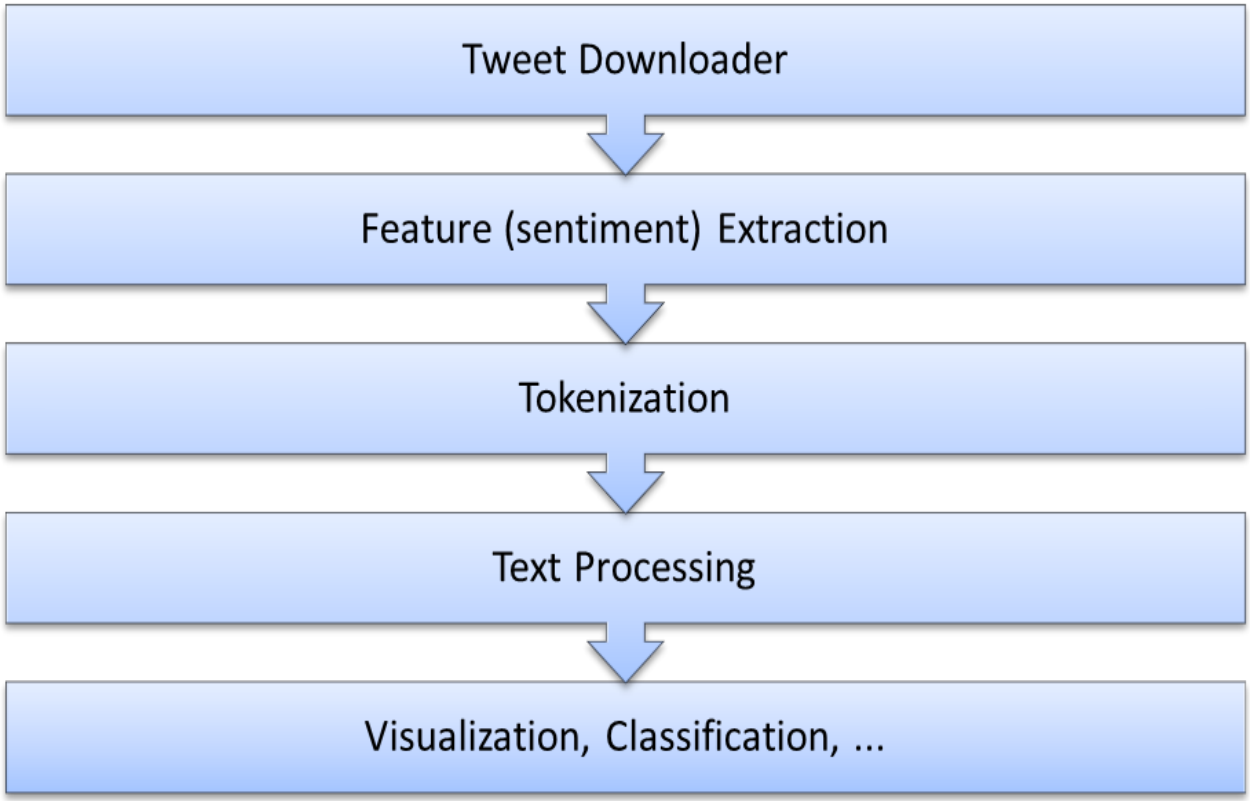
- Basic tasks:
  - Determine polarity (positive/negative/neutral)
  - Determine subjectivity/objectivity

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

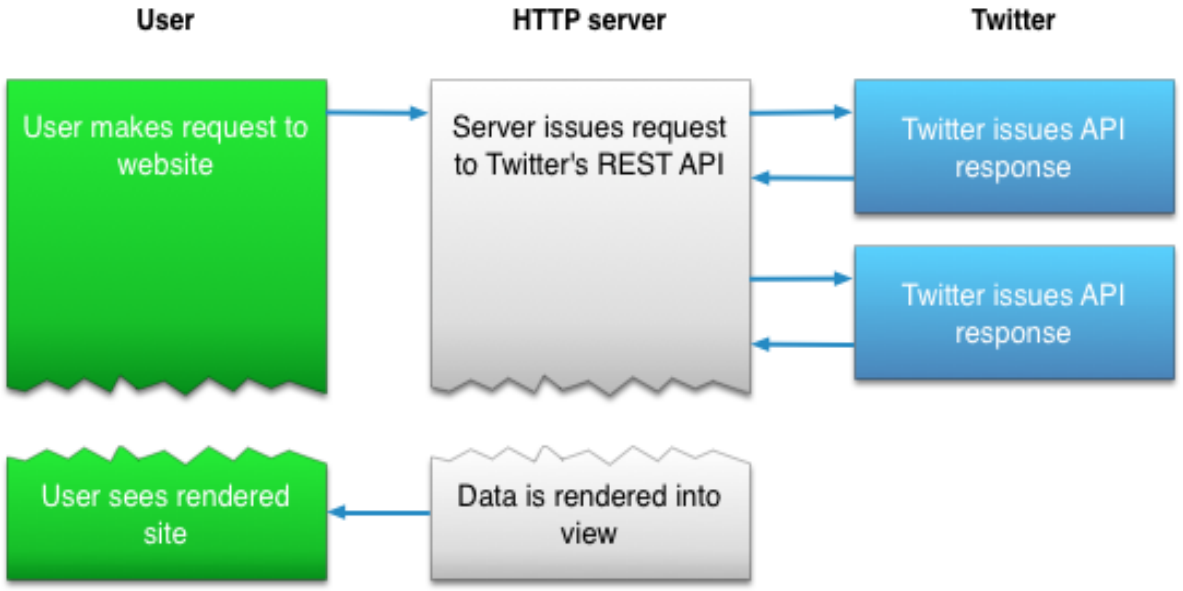
Opinion Mining in Different Industries

- 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 in-person 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
Feature (sentiment) Extraction
Tokenization
Text Processing
Visualization, Classification, ...

Tweet Downloader (through Twitter API)



source: developer.twitter.com

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 SentiWordNet 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)

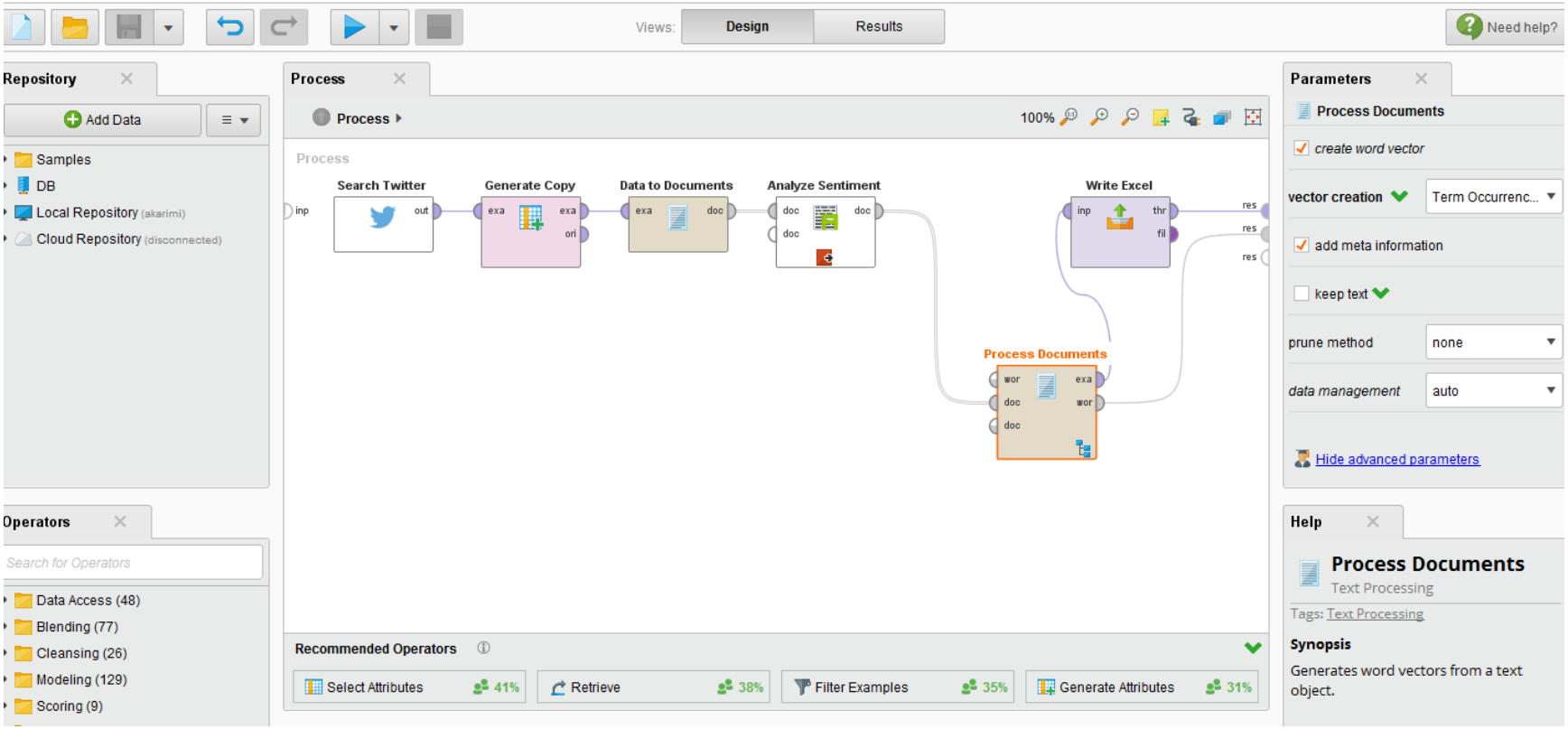
Challenges of Sentiment Extraction

- 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

RapidMiner Demo.



Output of ‘Search Twitter’ operator

ExampleSet (100 examples, 1 special attribute, 11 regular attributes)

Filter (100 / 100 examples): all

Row No.	Id	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text
1	9235611625...	Oct 26, 2017 ...	Special Olym...	14496153	?	-1	en	<a href="http:/...	#TBT: Action from the 2003 Fall Games
2	9235586638...	Oct 26, 2017 ...	Titan Shops/...	15864633	?	-1	en	<a href="http:/...	Get the latest CSUF gear, like the CSU
3	9235573557...	Oct 26, 2017 ...	Democratic P...	21231110	?	-1	en	<a href="http:/...	Spend this Halloween spooking away
4	9235561113...	Oct 26, 2017 ...	SETC	115578122	?	-1	en	<a href="http:/...	Meet graduate #sounddesign program
5	9235529879...	Oct 26, 2017 ...	Zesty MLB Do...	7120219227...	?	-1	en	<a href="http:/...	Wild night for Astros bullpen, but forme
6	9235489649...	Oct 26, 2017 ...	Sarah T. Rob...	23451153	?	-1	en	<a href="http:/...	Hey, remember how the interim dean a
7	9235447612...	Oct 26, 2017 ...	Matt Rhodes	244365222	?	-1	en	<a href="http...	Former Cal State Fullerton player Chris
8	9235408270...	Oct 26, 2017 ...	Joseph Rex	408119751	NCAACWS	41404125	en	<a href="http:/...	@NCAACWS @BaseballTitans @csuf
9	9235324888...	Oct 26, 2017 ...	R. W. Salatto	546627481	?	-1	en	<a href="http:/...	Finally a titan I want to support. Boooo
10	9235155648...	Oct 26, 2017 ...	Ron Bohning	32693541	?	-1	en	<a href="http...	Former #Cal_State_Fullerton player Ch
11	9235034918...	Oct 26, 2017 ...	MLB NewzBot	613609261	?	-1	en	<a href="http:/...	Yahoo! Sports - https://t.co/94KUaN5oc
12	9234899243...	Oct 26, 2017 ...	Robert Cerra	86503703	PoliceFound	2360426665	en	<a href="http:/...	@PoliceFound @csuf great idea - I was
13	9234873503...	Oct 26, 2017 ...	Fresh Squee...	8305865404...	?	-1	en	<a href="http...	Former Cal State Fullerton player Chris
14	9234857949...	Oct 26, 2017 ...	Ocduilaw.com	69956087	?	-1	en	<a href="http:/...	https://t.co/uY7RFRFxEz - Former Cal S
15	9234788225...	Oct 26, 2017 ...	Christine Mikt...	4813863308	fresnophile	9107175363...	en	<a href="http:/...	@fresnophile @Under_Erasure @Hen
16	9234719350...	Oct 26, 2017 ...	Lakers Fans	3038730086	?	-1	en	<a href="http...	#LosAngelesLakers #Lakers #LAL For
17	9234593632...	Oct 26, 2017 ...	Richard Krau...	7143860618...	?	-1	en	<a href="http:/...	#USA #LAT #NEWS broadast by #EME
18	9234588115...	Oct 26, 2017 ...	Football Field	106277663	?	-1	en	<a href="http:/...	Wild night for Astros' bullpen, but forme

Output of ‘Tokenize’ operator

Final Word list (no filtering)

Word	Attribute Name	Total Occurences	Document Occurences ↓
Cal	Cal	51	51
Fullerton	Fullerton	52	51
State	State	51	51
for	for	51	51
the	the	55	51
Devenski	Devenski	44	44
gets	gets	42	42
player	player	41	41
Astros	Astros	40	40
win	win	40	40
night	night	39	39
on	on	33	31
Chris	Chris	26	26
to	to	30	26
Former	Former	23	23
but	but	23	23
wild	wild	21	21
Wild	Wild	20	20
bullpen	bullpen	20	20

Next…

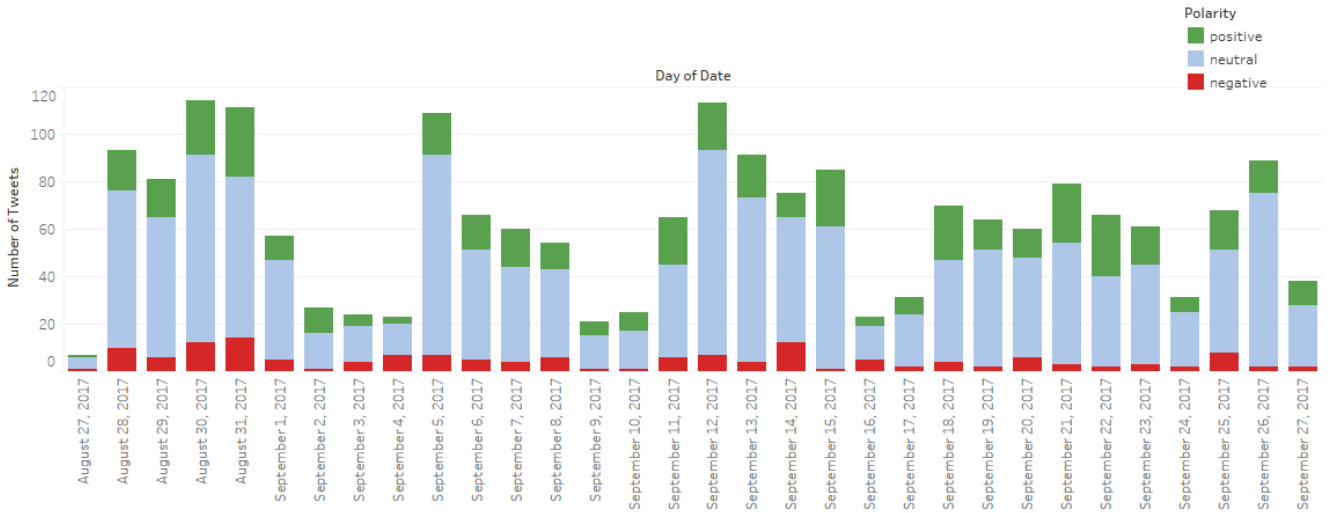
- Visualization
  - Format the data
  - Build Tableau dashboards at both Tweet and word levels

- 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

of Tweets per Day

Twitter Analysis

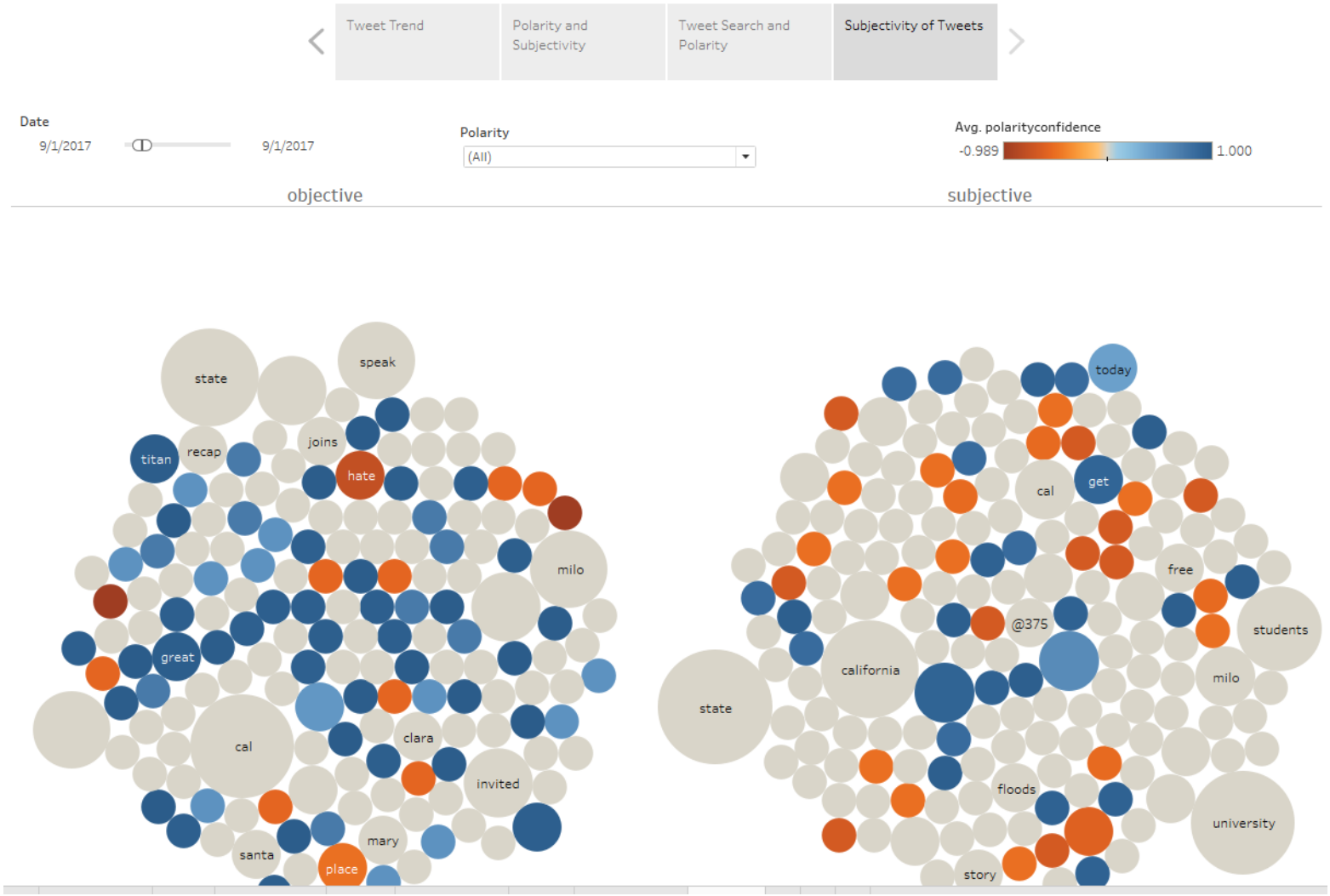
Tweet Trend | Polarity and Subjectivity | Tweet Search and Polarity | Subjectivity of Tweets



From User	Text2	
1 Los Angeles	Los Angeles News Search LA (2 Cal State Fullerton students diagnosed with chicken pox has students, staff ...) - https://t.co/voqzACDcHO	0.0
	Los Angeles News Search LA (Controversial speaker Milo Yiannopoulos invited to lecture at CSU Fullerton) 1 ... -... https://t.co/KW2zrOISjF	0.0
	Los Angeles News Search LA (Lifelong learning brings seniors to Cal State Fullerton campus) 1 Los Angeles - https://t.co/7sYwGHAkzH	0.0
49er Shops Inc.	This portable, waterproof, Bluetooth #JBL speaker at #BeachTech has surprisingly powerful sound that'll make CSUF a... https://t.co/1DVppadduq	1.0
__Luke__( ):	@Beth_CSUF "Lack of corporate responsibility" I love that phrase. 10/10	0.0
	@Beth_CSUF topic for discussion tonight: the appropriate number of wrong password attempts before you get locked out of your email ??	0.0
	@csuf @MeganA101 y'all aren't loyal and you like Megan more than me	0.0
	@csuf do we need to bring printed tickets for concert tonight or can we scan on our phones? ??	0.0
	I love daddy newsom ?? https://t.co/nKaiTpyFTQ	0.0
	I really can't get over how good Millie looks tonight. Nothing but respect for MY president @csuf	0.0
	Leave it to Advancement to schedule concert under the stars on the day the world is ending #TitansReachHigher @csuf https://t.co/aEkdRRsAiW	1.0
	Okay PMG looks so good! Damn girl you work those Diamonds™!!! https://t.co/W2dPCeWk0w	0.0
-a	I'm at California State University Fullerton - @csuf in Fullerton, CA https://t.co/GKXGREJfsn	0.0
:D Los Angeles	Milo Yiannopoulos plans to speak at Cal State Fullerton despite protests and criticism https://t.co/rwusHw3IVG	0.0

Polarity & Subjectivity

Subjectivity of Tweets



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