

Be Smart Twitter



Zihan Zhou
Nicholas Montoya
Ziyang Yang

Questions

- How does people's mood change according to the time period of a day? Are any words associated with this shift?
- How do people use the Emojis/Emoticons (':-') to express their real sentiments? What are some downfalls in current language processing techniques regarding this and what could improve them?

Data Preparation Work

- Removing usernames
- Removing miscellaneous characters
- Tokenizing tweets
 - Vectorizing tweets
- Removing duplicate tweets from various data sources
- Separate specific times into ranges

	EmotionLevel	TwitterID	PostTime	IsQuery	Username	Content
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....
5	0	1467811372	Mon Apr 06 22:20:00 PDT 2009	NO_QUERY	joy_wolf	@Kwesidei not the whole crew
6	0	1467811592	Mon Apr 06 22:20:03 PDT 2009	NO_QUERY	mybirch	Need a hug
7	0	1467811594	Mon Apr 06 22:20:03 PDT 2009	NO_QUERY	coZZ	@LOLTrish hey long time no see! Yes.. Rains a...
8	0	1467811795	Mon Apr 06 22:20:05 PDT 2009	NO_QUERY	2Hood4Hollywood	@Tatiana_K nope they didn't have it
9	0	1467812025	Mon Apr 06 22:20:09 PDT 2009	NO_QUERY	mimismo	@twittera que me muera ?
10	0	1467812416	Mon Apr 06 22:20:16 PDT 2009	NO_QUERY	erinx3leannexo	spring break in plain city... it's snowing
11	0	1467812579	Mon Apr 06 22:20:17 PDT 2009	NO_QUERY	pardonlauren	I just re-pierced my ears

Tools used

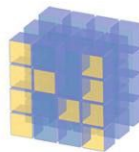
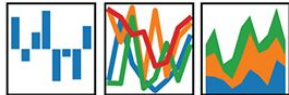
- Natural Learning Tool Kit (NLTK)
- NumPy
- Pandas
- matplotlib



Natural Language Analysis
with Python NLTK

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



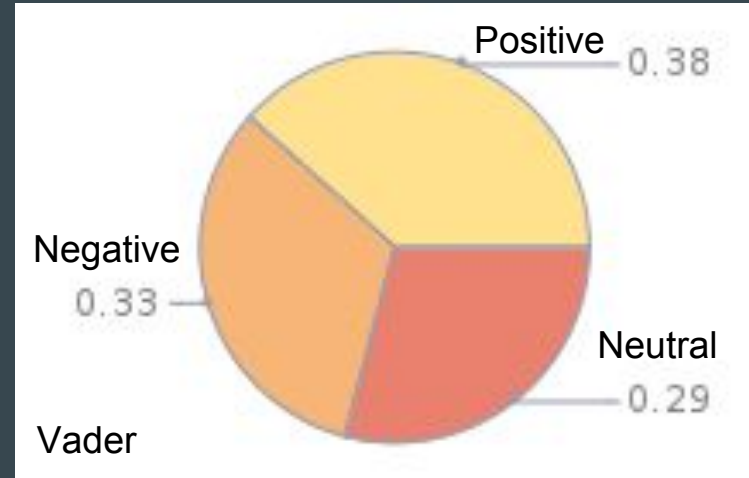
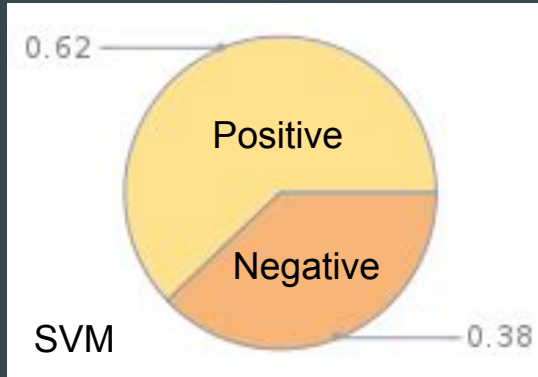
NumPy



matplotlib

Classification & Clustering

- Made SVM to classify tweets
- Classified tweets as positive, negative, or neutral along with subjective vs. objective using Vader (NLTK)
- K-Means clustering with various 'K'



Knowledge Gained

- Positive sentiment occurrence peaks in the morning
- Negative sentiment is relatively constant throughout the day
- Neutral sentiment peaks during daytime hours but intensity of others rises, surprisingly high amount of URLs.
- Sentiment around shopping changes from morning to night
- Emojis/Emoticons are easily misinterpreted due to irony and sarcasm
- It's more possible that people use negative words in negative sentiment than using the positive words in positive sentiment

Application

- Target advertisements between 7:30-9pm
- Release top headlines during standard work hours
- Improve language processing libraries
- Collate and classify reviews (like | dislike | neutral)
- Analyze and predict political elections

Thank you