Data & Text Analysis

Sentiment Analysis

IMT 547 - Social Media Data Mining and Analysis

9-Feb-2021 (Week 6, Day 11)

Today’s Topics

- Sentiment Analysis
- Rule-based approach
- VADER
- LIWC
- TextBlob
- Empath - embedding-based approach
- Lab — sentiment
- Representing data for processing: Document Term Matrix
- Lab2 - data cleaning + sentiment

What is Sentiment Analysis

Examples: Positive or negative movie review?

Unbelievably disappointing

Richly applied satire, great plot twists.

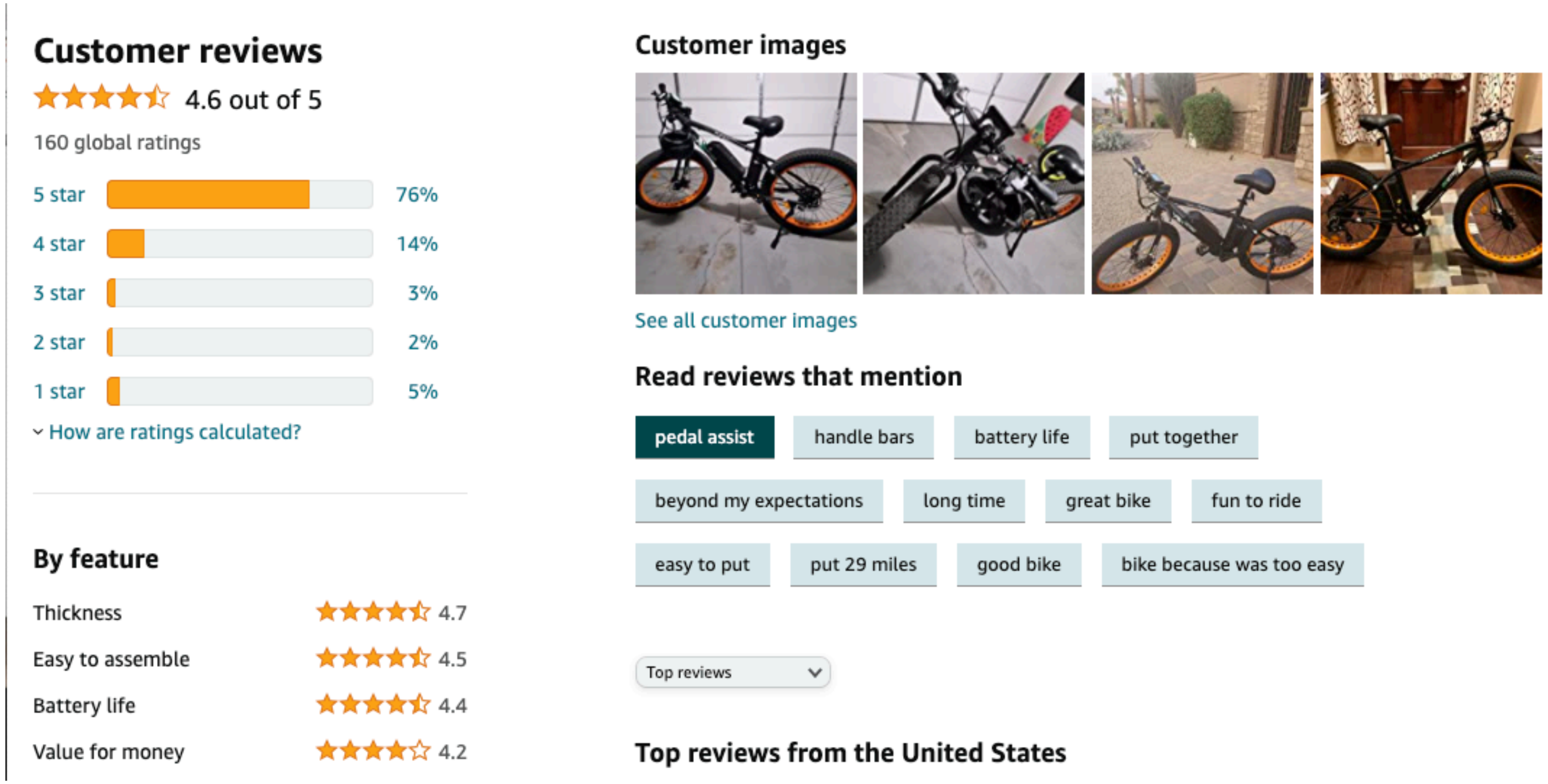
This is the greatest screwball comedy ever filmed

It was pathetic. The worst part about it was the boxing scenes

Text Classification — Sentiment Analysis

Sentiment analysis

Aspect detection in product reviews and measuring sentiment for each aspect



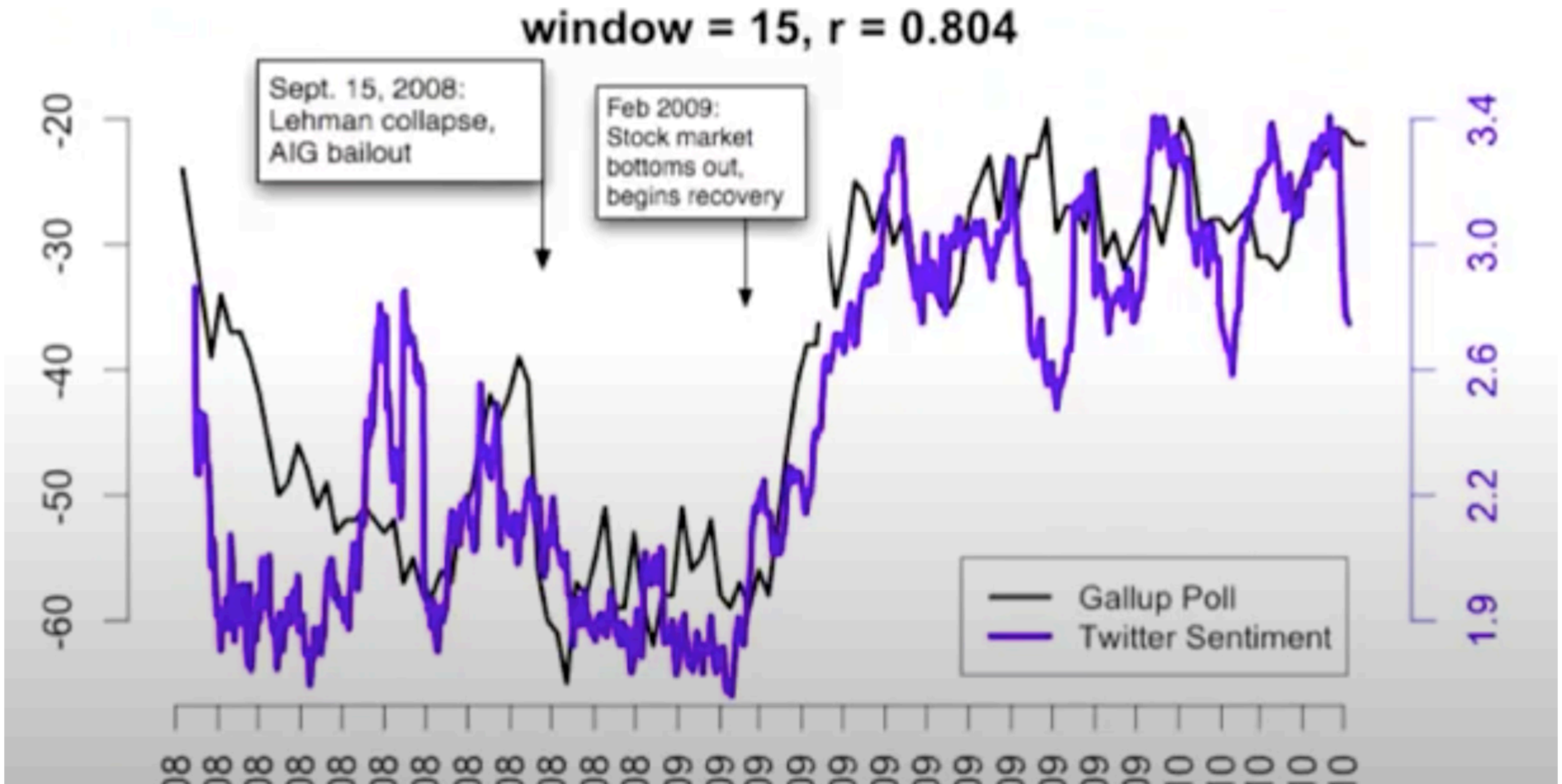
Extract aspects from the reviews, like *thickness*, *easy to assemble*, *battery life* and summarize from the various reviews the positive and negative sentiment

From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

Brendan O'Connor† Ramnath Balasubramanyan† Bryan R. Routledge§ Noah A. Smith†
brenocon@cs.cmu.edu rbalasub@cs.cmu.edu routledge@cmu.edu nasmith@cs.cmu.edu

http://www.cs.cmu.edu/~nasmith/papers/oconnor+balasubramanyan+routledge+smith.icwsm10.pdf

Sentiment ratio & Consumer Confidence survey



From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

Brendan O'Connor[†] **Ramnath Balasubramanyan**[†] **Bryan R. Routledge**[§] **Noah A. Smith**[†]
brenocon@cs.cmu.edu rbalasub@cs.cmu.edu routledge@cmu.edu nasmith@cs.cmu.edu

http://www.cs.cmu.edu/~nasmith/papers/oconnor+balasubramanyan+routledge+smith.icwsm10.pdf

Sentiment ratio & Consumer Confidence survey

Message Retrieval

We only use messages containing a topic keyword, manually specified for each poll:

- For consumer confidence, we use *economy*, *job*, and *jobs*.
- For presidential approval, we use *obama*.
- For elections, we use *obama* and *mccain*.

Derive day-to-day sentiment scores by counting positive and negative messages.

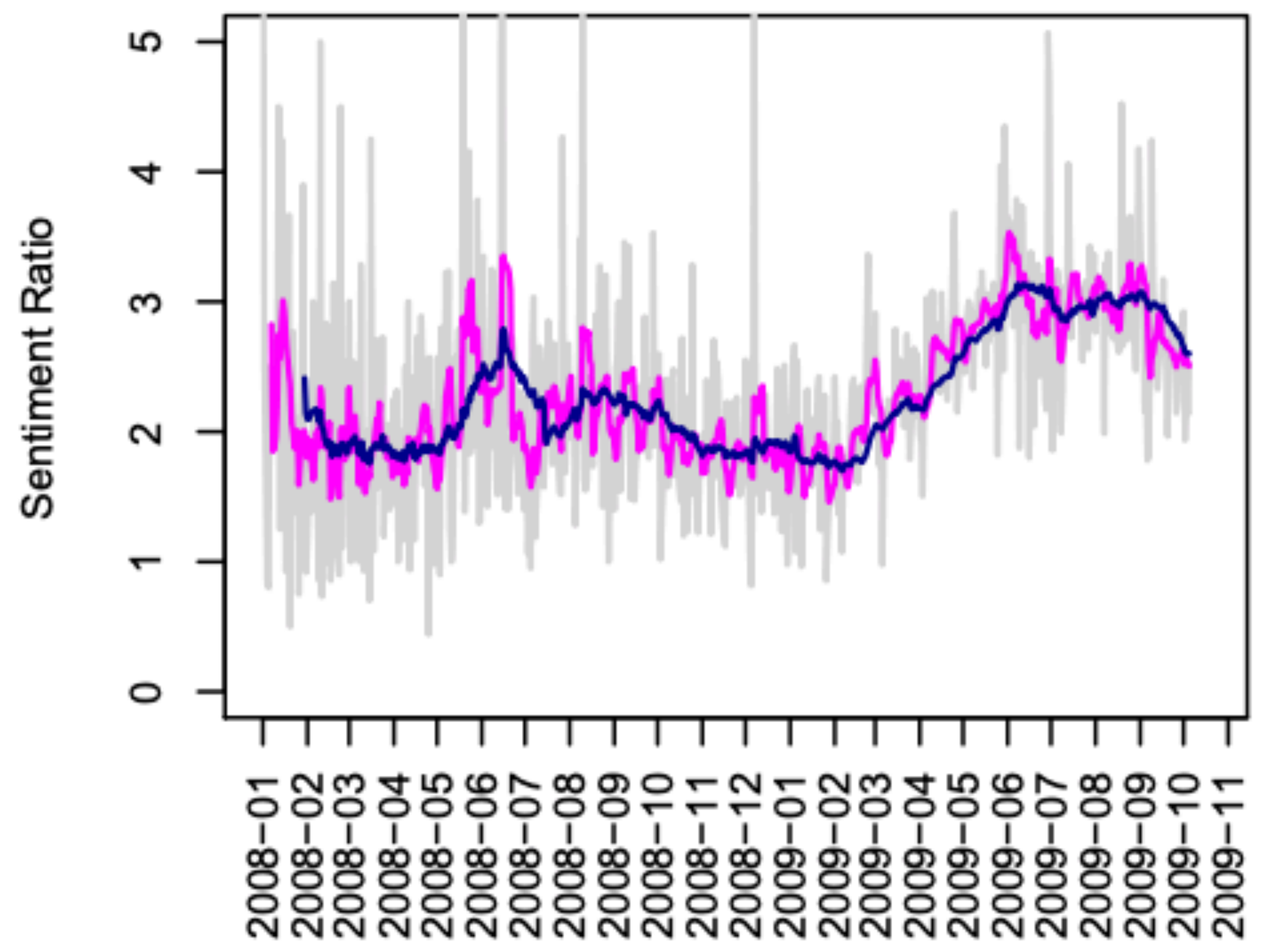
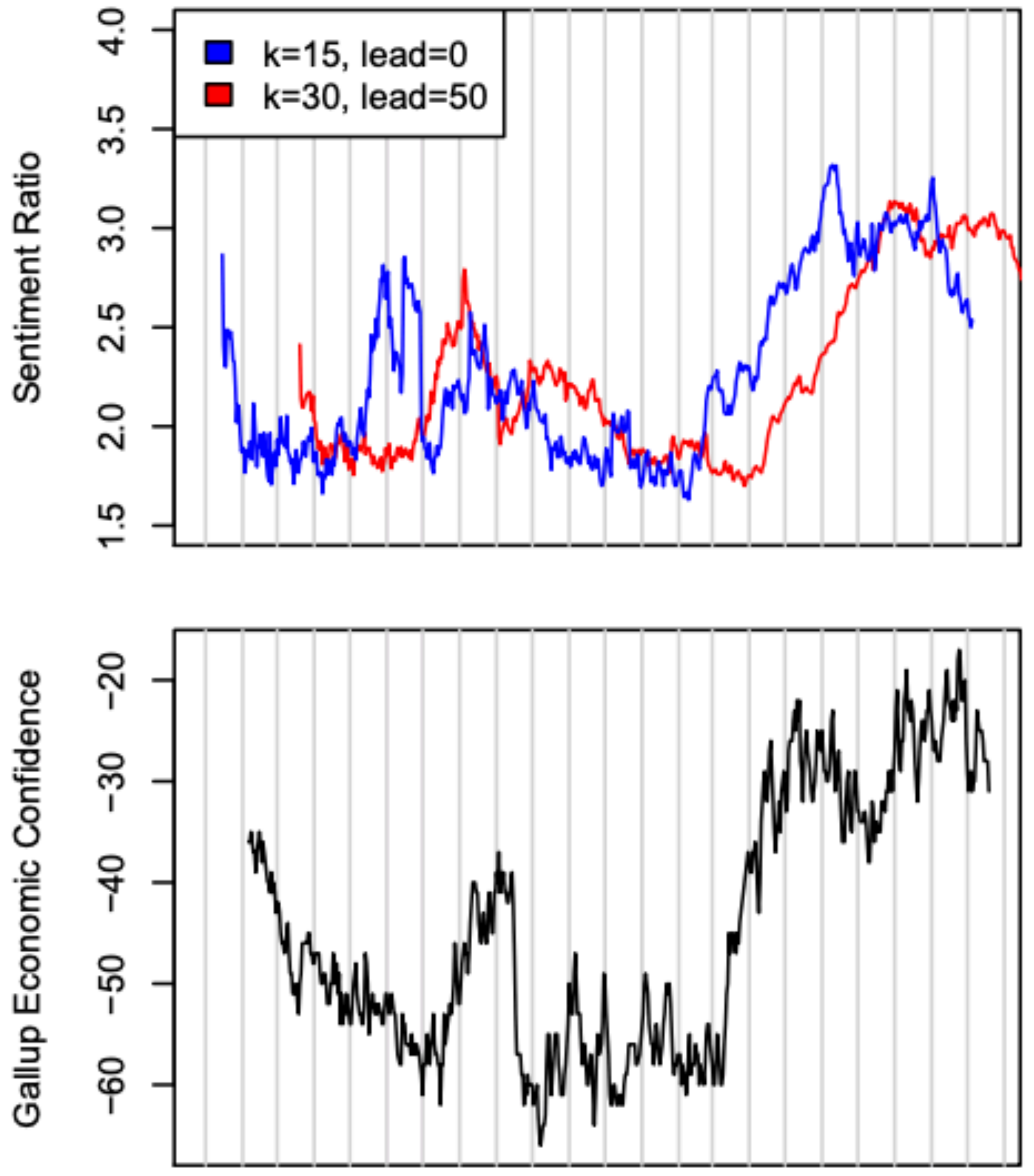


Figure 5: Moving average MA_t of sentiment ratio for *jobs*, under different windows $k \in \{1, 7, 30\}$: no smoothing (gray), past week (magenta), and past month (blue). The unsmoothed version spikes as high as 10, omitted for space.

jobs sentiment ratio compared to the measures of consumer confidence and Gallup Daily polls



Sentiment - more fine grained (anxiety index)

Widespread Worry and the Stock Market

Eric Gilbert and Karrie Karahalios
Department of Computer Science
University of Illinois at Urbana-Champaign
[egilber2, kkarahal]@cs.uiuc.edu

Twitter mood predicts the stock market

Johan Bollen[a,*,1], Huina Mao[a,1], Xiaojun Zeng[b]
[a] School of Informatics and Computing, Indiana University, 919 E. 10th Street, Bloomington, IN 47408, United States
[b] School of Computer Science, University of Manchester, Kilburn Building, Oxford Road, Manchester M13 9PL, United Kingdom

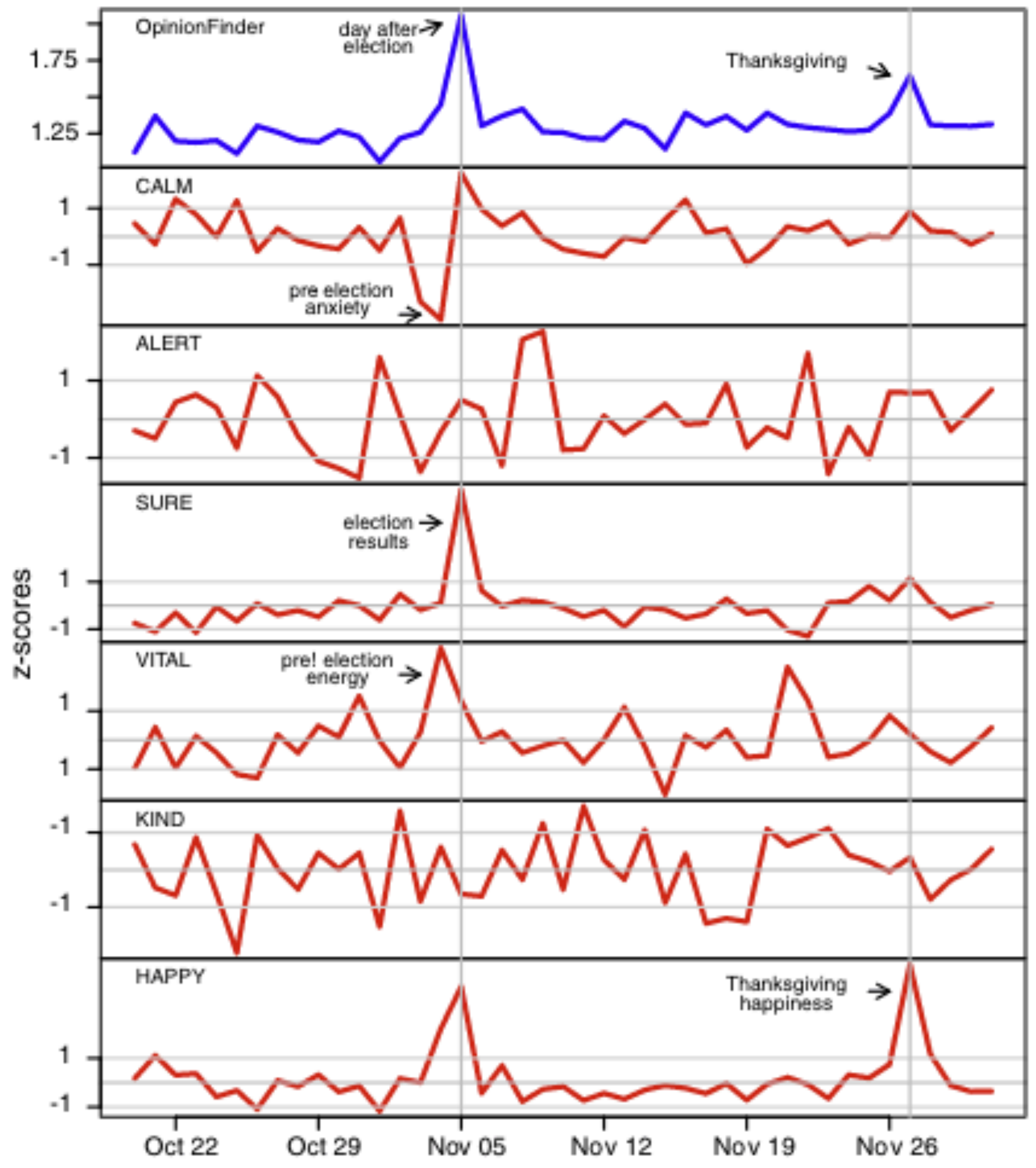


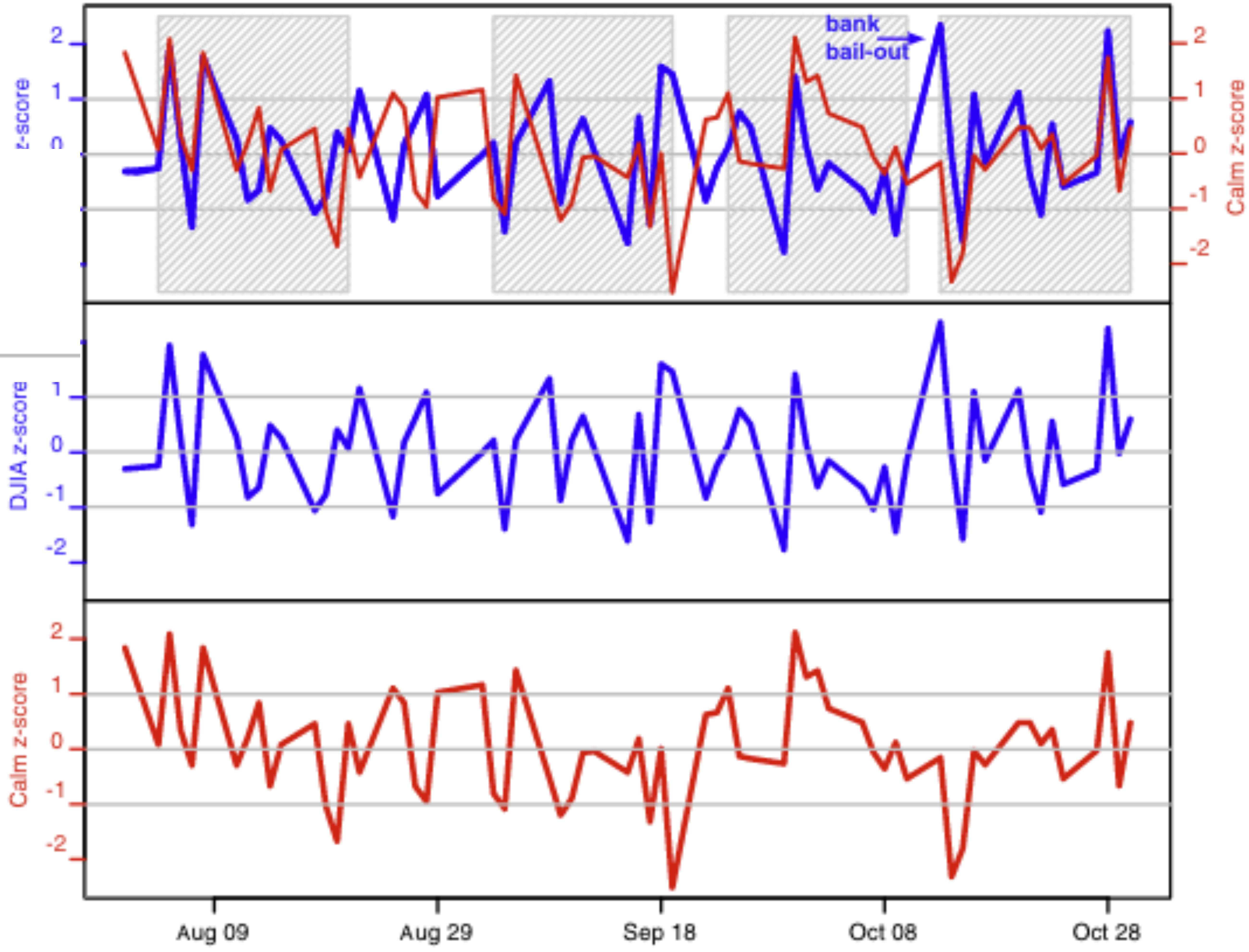
Fig. 2. Tracking public mood states from tweets posted between October 2008 to December 2008 shows public responses to presidential election and thanksgiving.

Sentiment & Stock market

Twitter mood predicts the stock market

Johan Bollen [a,*,1], Huina Mao [a,1], Xiaojun Zeng [b]

[a] School of Informatics and Computing, Indiana University, 919 E. 10th Street, Bloomington, IN 47408, United States
[b] School of Computer Science, University of Manchester, Kilburn Building, Oxford Road, Manchester M13 9PL, United Kingdom



CALM predicts Dow Jones Industrial Average (DJIA) 3-days later.

What is Sentiment Analysis

Many difference names & used for various tasks

- Sentiment mining
- Opinion extraction
- Opinion mining
- Subjectivity analysis

- Movies: +ve or -ve reviews
- Products: what do people think of this new iPhone?
- Public sentiment: Consumer confidence (replacing gallup poll)
- Politics: People think about this candidate or issue?
- Personal analytics: What are people's view about X? Are people happy or sad — *happiness index*?
- Prediction: predict election outcomes or market trends from sentiment

Sentiment and Affective States

Scherer typology of affective states

- **Emotion**: brief organically synchronized. Often relevant in evaluation of a major event. e.g.: *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling. e.g: *Cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affected stance toward another person in a in a specific interaction, coloring the interpersonal exchange in that situation (*distant, cold, warm, supportive, contemptuous*)
- **Attitude**: relatively enduring, affectively colored belief, preference, or predisposition towards objects or persons (*liking, loving, hating, valueing, desiring*)

Sentiment is often a measure of attitudes

- **Personality trait**: emotionally laden, stable personality dispositions and behavior tendencies (*nervous, anxious, reckless, morose, hostile, envious, jealous*)

Ranganath, Rajesh, Dan Jurafsky, and Daniel A. McFarland. **"Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates**.” Computer Speech & Language 27.1 (2013): 89-115.

Sentiment Analysis

Sentiment Analysis is the detection of attitudes

1. **Holder** (source) of attitude: Who has the attitude
2. **Target** (aspect) : what is it that we have this attitude about or about whom
3. **Type** of attitude:
   - From a set of types: e.g. *like, love, hate, value, desire*
   - Or more commonly simple weighted polarity - *positive, negative, neutral*
4. **Text** containing the attitude: e.g., the sentence or the entire document (tweets, reddit posts, news article etc.)

Sentiment Analysis

Sentiment Analysis is the detection of attitudes

- **Simplest task:** Is the attitude of this text positive or negative?
- **More complex:** Rank the attitude of this text from 1 to 5
- **Advanced**: Detect the ***target***, ***source***, or ***complex attitude types***

Sentiment and Affective States - *LIWC can come in handy*

Scherer typology of affective states

- **Emotion**: brief organically synchronized. Often relevant in evaluation of a major event. e.g.: *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling. e.g: *Cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affected stance toward another person in a in a specific interaction, coloring the interpersonal exchange in that situation (*distant, cold, warm, supportive, contemptuous*)
- **Attitude**: relatively enduring, affectively colored belief, preference, or predisposition towards objects or persons (*liking, loving, hating, valueing, desiring*)

Sentiment is often a measure of attitudes

- **Personality trait**: emotionally laden, stable personality dispositions and behavior tendencies (*nervous, anxious, reckless, morose, hostile, envious, jealous*)

Ranganath, Rajesh, Dan Jurafsky, and Daniel A. McFarland. **"Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates.**" Computer Speech & Language 27.1 (2013): 89-115.

LIWC

Take a look at LIWC dictionary and think which dimension can be useful for your project

Dictionary of words representing psychological dimensions

Multiple dimensions: https://liwc.wpengine.com/compare-dictionaries/

52 Cause	53 Discrep	54 Tentat		55 Certain	56 Differ
activat*	abnormal*	allot	undecided*	absolute	actually
affect	besides	almost	undetermin*	absolutely	adjust*
affected	could	alot	unknowing	accura*	against
affecting	could've	ambigu*	unknowingly	all	ain't
affects	couldn't	any	unknown	altogether	aint
aggravat*	couldnt	anybod*	unlikel*	always	alternativ*
allow*	couldve	anyhow	unresolv*	apparent	although
attribut*	desir*	anyone*	unsettl*	assur*	apart
based	expect*	anything	unsure*	blatant*	aren't
basis	hope	anytime	usually	certain*	arent
bc	hoped	anywhere	vague	clear	but
because	hopeful	apparently	vaguely	clearly	can't
bosses	hopefully	appear	vagueness	commit	cannot
caus*	hopes	appeared	vaguer	commitment*	cant
change	hoping	appearing	vaguest	commits	despite
changed	ideal*	appears	variab*	committed	didn't
changes	if	apprehens*	varies	committing	didnt
changing	impossible	approximat*	vary	complete	differ
compel*	inadequa*	arbitrar*	virtually	completed	differed
compliance	lack	assum*	wonder	completely	difference*
compliant	lacked	barely	wondered	completes	different
complied	lacking	bet	wondering	confidence	differential
complies	lacks	bets	wonders	confident	differentiat*

BB	BC	BD	BE	BF	BG	BH
32 Negemo			33 Anx	34 Anger		35 Sad
ignorant	poorest	tragic	afraid	abuse*	poison*	abandon*
ignore	poorly	trauma*	alarm*	abusi*	prejudic*	agoniz*
ignored	poorness*	trembl*	anxiety	aggravat*	prick*	agony
ignores	powerless*	trick	anxious	aggress	protest	alone
ignoring	prejudic*	tricked	anxiously	aggressed	protested	bereave*
immoral*	pressur*	trickier	anxiousness	aggresses	protesting	broke
impatien*	prick*	trickiest	apprehens*	aggressing	protests	cried
impersonal	problem*	tricks	asham*	aggression*	punish*	cries
impolite*	protest	tricky	aversi*	aggressive	pushy	crushed
inadequa*	protested	trite	avoid*	aggressively	rage*	cry
incompeten*	protesting	trivial	awkward	aggressor*	raging	crying
indecis*	protests	troubl*	confuse	agitat*	rape*	depress*
ineffect*	puk*	turmoil	confused	anger*	raping	depriv*
inferior	punish*	twitchy	confusedly	angrier	rapist*	despair*
inferiority	pushy	ugh	confusing	angriest	rebel*	devastat*
inhibit*	queas*	uglier	desperat*	angry	resent*	disappoint*
insecur*	rage*	ugliest	discomfort*	annoy	revenge*	discourag*
insincer*	raging	ugly	distraught	annoyed	ridicul*	dishearten*
insult*	rancid*	unaccept*	distress*	annoying	rude	disillusion*
interrup*	rape*	unattractive	disturb*	annoys	rudely	doom*
intimidat*	raping	uncertain*	doubt*	antagoni*	sarcas*	dull
irrational*	rapist*	uncomfortabl*	dread*	argh*	savage*	emptier
irrita*	rebel*	uncontrol*	dwell*	argu*	sceptic*	emptiest

Tausczik, Yla R., and James W. Pennebaker. "**The psychological meaning of words: LIWC and computerized text analysis methods**." *Journal of language and social psychology* 29.1 (2010): 24-54.

VADER

Rule-based sentiment analysis

“We incorporated consideration for rule 4 by splitting the text into segments around the contrastive conjunction “*but*”, and diminished the total sentiment intensity of the text preceding the conjunction by 50% while increasing the sentiment intensity of the post-conjunction text by 50%.”

Lots of human-labeling: Over 9,000 token features were rated on a scale from "[–4] Extremely Negative" to "[4] Extremely Positive", with allowance for "[0] Neutral (or Neither, N/A)".

TextBlob

Another rule-based sentiment analyzer

https://textblob.readthedocs.io/en/dev/quickstart.html

Powered by rules and category annotations


```
<word form="abhorrent" wordnet_id="a-1625063" pos="JJ" sense="offensive to the mind" polarity="-0.7" subjectivity="0.8" intensity="1.0" r
<word form="able" cornetto_synset_id="n_a-534450" wordnet_id="a-01017439" pos="JJ" sense="having a strong healthy body" polarity="0.5" su
<word form="able" wordnet_id="a-00001740" pos="JJ" sense="(usually followed by 'to') having the necessary means or skill or know-how or a
<word form="able" wordnet_id="a-00306663" pos="JJ" sense="having inherent physical or mental ability or capacity" polarity="0.5" subjecti
<word form="able" wordnet_id="a-00510348" pos="JJ" sense="have the skills and qualifications to do things well" polarity="0.5" subjectivi
<word form="above" cornetto_synset_id="n_a-504850" wordnet_id="a-00125993" pos="JJ" sense="appearing earlier in the same text" polarity="
<word form="abridged" cornetto_synset_id="d_a-9176" wordnet_id="a-00004413" pos="JJ" sense="(used of texts) shortened by condensing or re
<word form="abrupt" cornetto_synset_id="n_a-505100" wordnet_id="a-00640520" pos="JJ" sense="surprisingly and unceremoniously brusque in m
<word form="abrupt" cornetto_synset_id="n_a-529169" wordnet_id="a-01145151" pos="JJ" sense="extremely steep" polarity="0.0" subjectivity=
<word form="abrupt" wordnet_id="a-01143585" pos="JJ" sense="exceedingly sudden and unexpected" polarity="0.0" subjectivity="1.0" intensit
<word form="abrupt" wordnet_id="a-02294122" pos="JJ" sense="marked by sudden changes in subject and sharp transitions" polarity="0.0" sub
```


Word net: lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations.

https://wordnet.princeton.edu/

Deep learning vs. Simple rule-based VADER

Comparing with more sophisticated models

Recursive Deep Models for Semantic Compositionality Over Sentiment Treebank, *EMNLP-2013*

Vs.

VADER - *ICWS 2014*

"…for simple binary (positive/negative) classification on single sentences is around 80%, and that for the more difficult multiclass case that includes a third (neutral) class, accuracies tend to hover in the 60% range for social media text.

VADER's simple rule-based approach are on par with such sophisticated benchmarks "

Advantages of the simplicity of VADER

- quick and computationally economical without sacrificing accuracy.
- the lexicon and rules used by VADER are directly accessible, not hidden within a machine-access- only black-box. VADER is therefore easily inspected, un- derstood, extended or modified
- VADER is domain agnostic – it does not require an extensive set of training data, yet it performs well in diverse domains.

How many lexicons were discussed in the paper?

Embedding-based approach

EMPATH -

http://empath.stanford.edu/

https://github.com/Ejhfast/empath-client

Empath is a tool for analyzing text across lexical categories (similar to LIWC), and also generating new lexical categories to use for an analysis.

python library:

pip install empath

EMPATH - http://empath.stanford.edu/

Empath

Empath can generate new lexical categories and analyze text over 200 built-in human-validated categories. This is a web demo; you can download our python tool to more easily run larger analyses, or take a look at our recent CHI paper.

Generate category:

spoon, oven, counter

Generate

pancake, fridge, oatmeal, pan, waffle, cutting_board, burner, stove, ladle, coffee_pot, kitchen_counter, blender, pancake_mix, freezer, pastry, mugs, oven, dishwasher, donut, small_bowl, paper_plate, mug, frying_pan, glass_bowl, syrup, cabinet, plastic_spoon, table, glass_plate, whisk, coffee_machine, kettle, kitchen_sink, microwave, cereal_box, marble_counter, mashed_potatoes, spatula, lasagna, wooden_spoon, container, empty_cup, kitchen, bagel, batter, silver_tray, fork, plastic_container, bowl, bowel, pantry, silverware, pasta, skillet, refrigerator, cereal_bowl, spoon, kitchen_bench, oven_mitts, empty_bowl, last_plate, platter, counter_top, island_counter, brownie, banana, big_bowl, plate, large_bowl, coffee_table, toaster, carton, pizza_box, pot, kitchen_island, empty_plate, breakfast_bar, containers, sink, dish, cupboard, bowls, coffee_maker, coffee_mug, teapot, cooker, jug, noodles, cutlery, refridgerator, own_plate, counter, chopsticks, pretzel, big_plate, omelette, tray, countertop, plates, small_plate

Analyze text over default categories

he hit the other person

Analyze

negative_emotion	1
violence	1
pain	1
movement	1

Lab 1

BREAK

Be back at 9:45am

Quantitative Methods | *Text Classification*

Examples: Positive or negative movie review?

	Labels
Unbelievably disappointing. Worst movie ever.	0
Richly applied satire, great plot twists.	1
This is such a great comedy	1
It was pathetic. The worst part about it was the boxing scenes	0

INPUT: document *d*
set of classes C={c1, c2,…},
labeled documents (d1,c1), (d2, c1), (d3, c2),……

OUTPUT: learned classifier. *Given a new document classifier can predict the class*

Adapted from Jurafsky's NLP lectures

After data cleaning: # Representing Data in Document Term Matrix

Rows: Data or instance or sample or records

Columns: Features or predictors or independent variable or attributes (X)

Special column - Target or outcome or response or dependent variable (Y)

Computer can understand this matrix format

	X							
Tokens	***"great"***	***"worst"***	***"plot"***	**....**	**word count**	**reviewer reputation**	**...**	**Y**
Unbelievably disappointing. Worst worst movie ever.		2						0
Richly applied satire, great plot twists.								1
This is such a great comedy								1
It was pathetic. The worst part about it was the boxing scenes								0

Lab2

Few announcements!

NEXT CLASS: Please show up to work in groups

PS2: Will be released today. Start when released.

All project pitches are now available for others to see:

https://canvas.uw.edu/courses/1434897/files/folder/Projects/Project-pitch-submissions