

Midterm Elections in the Age of Big Data

Sarah Shugars, Network Science Institute



Northeastern University
Network Science Institute

Northeastern University
College of Social Sciences and Humanities

Case Study: Midterm Impact of Kavanaugh Hearing



NYTIMES.COM

Here Are the 5 Senators Who Will Decide

Kavanaugh's Fate

Save

<https://www.nytimes.com/2018/10/02/us/politics/kavanaugh-undecided-senators.html>

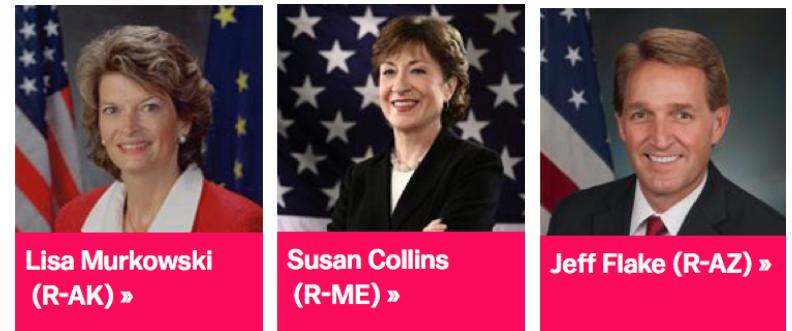
Case Study: Midterm Impact of Kavanaugh Hearing



NYTIMES.COM

Here Are the 5 Senators Who Will Decide Kavanaugh's Fate

Save



<https://www.nytimes.com/2018/10/02/us/politics/kavanaugh-undecided-senators.html>

<https://www.vox.com/2018/7/12/17561692/senate-kavanaugh-confirmation-tracker>

Case Study: Midterm Impact of Kavanaugh Hearing

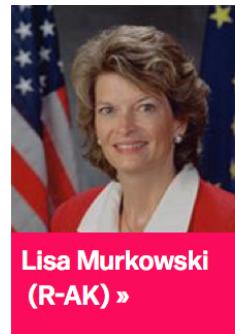


THEHILL.COM

Heitkamp to vote 'no' on Kavanaugh

Democratic Sen. Heidi Heitkamp (N.D.) said on Thursday that...

Save



**Lisa Murkowski
(R-AK) »**



**Susan Collins
(R-ME) »**



Jeff Flake (R-AZ) »



**Joe Manchin
(D-WV) »**



**Heidi Heitkamp
(D-ND) »**

[https://thehill.com/homenews/senate/
409948-heitkamp-to-vote-no-on-kavanaugh](https://thehill.com/homenews/senate/409948-heitkamp-to-vote-no-on-kavanaugh)

[https://www.vox.com/2018/7/12/17561692/
senate-kavanaugh-confirmation-tracker](https://www.vox.com/2018/7/12/17561692/senate-kavanaugh-confirmation-tracker)

Case Study: Midterm Impact of Kavanaugh Hearing



THEHILL.COM

Heitkamp to vote 'no' on Kavanaugh

Democratic Sen. Heidi Heitkamp (N.D.) said on Thursday that...

Save



Heitkamp's brother on her Kavanaugh vote: She may lose but she has to like who she sees in the mirror

Case Study: Midterm Impact of Kavanaugh Hearing

Heidi Heitkamp is fighting for her political life in North Dakota



By [Dana Bash](#) and Bridget Nolan, CNN



Heitkamp's brother on her Kavanaugh vote: She may lose but she has to like who she sees in the mirror

Case Study: Midterm Impact of Kavanaugh Hearing

Heidi Heitkamp is fighting for her political life in North Dakota

By [Dana Bash](#) and Bridget Nolan, CNN

The Washington Times

Sen. Heidi Heitkamp under pressure; 49 percent of North Dakota voters want to see her replaced: Poll



Heitkamp's brother on her Kavanaugh vote: She may lose but she has to like who she sees in the mirror

Case Study: Midterm Impact of Kavanaugh Hearing

**How is Senator Heitkamp's 'no'
vote on Kavanaugh's confirmation
being received?**

Case Study: Midterm Impact of Kavanaugh Hearing

**How is Senator Heitkamp's 'no'
vote on Kavanaugh's confirmation
being received?**

~OR~

Which narrative is dominating?

Scraping Tweets

Scraping Tweets



Scraping Tweets

The easy way:

<http://blogtrackers.fulton.asu.edu/>

<https://ifttt.com>

Scraping Tweets

The easy way:

<http://blogtrackers.fulton.asu.edu/>

<https://ifttt.com>

Pros:

- ...it is (relatively) easy.
- User interface with graphic search / export options

Scraping Tweets

The easy way:

<http://blogtrackers.fulton.asu.edu/>

<https://ifttt.com>

Pros:

- ...it is (relatively) easy.
- User interface with graphic search / export options

Cons:

- Limited capabilities
- Limited control

Tweet Tracker

<http://blogtrackers.fulton.asu.edu/>

Overview

Job Manager

Old Jobs

Tweetalyzer

Search/Export

Trends

TweetTube

Results

Tweets Images Videos

JSON XML

Export Selected Tweets Export All Tweets

Displaying 50 of 142 tweets

Load More Results Select All

Screen Name	Date/Time	Text	Follower Count	Lat/Lng	Source	Language	Retweet Count	Export
Yo64781220Young	Thu, 11 Oct 2018 04:02:38 GMT	RT @AriBerman: As @pemalevy reports, ND GOP wrote voter ID law after Heitkamp won her Senate election by 3,000 votes with strong Native sup...	25	N/A	🐦	en	0	<input type="checkbox"/>
Jockeyluis	Thu, 11 Oct 2018 04:05:57 GMT	RT @AriBerman: As @pemalevy reports, ND GOP wrote voter ID law after Heitkamp won her Senate election by 3,000 votes with strong Native sup...	675	N/A	🐦	en	0	<input type="checkbox"/>
kimmiejs	Thu, 11 Oct 2018 04:14:53 GMT	@MichaelAvenatti @BetoORourke Looks like Avenatti helped to win the election for Cruz. Good on ya Michael.	117	N/A	🐦	en	0	<input type="checkbox"/>
pugluvr555	Thu, 11 Oct 2018 04:17:50 GMT	@LDI_BruceWright I can only wish that the election was in my rear view mirror and Cruz was back on his way to Washi... https://t.co/Na2Uy7IEMV twitter.com/i/web/status/1...	163	N/A	🐦	en	0	<input type="checkbox"/>
witliftin	Thu, 11 Oct 2018 04:20:35 GMT	Given what ND has done to its Native population if Heitkamp loses that cannot be called a legitimate election. Same... https://t.co/WvP77INKFL twitter.com/i/web/status/1...	1579	N/A	🐦	en	0	<input type="checkbox"/>

Scraping Tweets

The hard way:

<https://developer.twitter.com/>

Scraping Tweets

The hard way:

<https://developer.twitter.com/>

Pros:

- More. Data.
- More flexibility: get tweets, friends/followers, reply chains

Scraping Tweets

The hard way:

<https://developer.twitter.com/>

Pros:

- More. Data.
- More flexibility: get tweets, friends/followers, reply chains

Cons:

- Requires some programming
- Requires being approved for a Twitter developer account

Twitter API

<https://developer.twitter.com/>

1. Register for a developer account & create an app

The screenshot shows the Twitter Developer portal interface. At the top, there's a purple navigation bar with links for 'Developer', 'Use cases', 'Products', 'Docs', and 'More'. Below the bar, the title 'Apps / Political Searches' is displayed. There are three tabs: 'App details' (which is active and highlighted in blue), 'Keys and tokens', and 'Permissions'. The main content area is titled 'App details' and contains sections for 'App icon', 'App Name', 'Description', 'Website URL', and 'Sign in with Twitter'. The 'App icon' section shows a small blue Twitter logo placeholder with the text 'App icon is default, click edit to upload.' The 'App Name' is listed as 'Political Searches'. The 'Description' is 'searching tweets with political keywords (eg, Trump, Clinton, etc)'. The 'Website URL' is 'http://www.networkscienceinstitute.org'. The 'Sign in with Twitter' section has the status 'Enabled'.

Twitter API

<https://developer.twitter.com/>

2. Write or find code

```
24     self.basePath=tilepath
25     os.system("mkdir -p %s"%(filepath))
26
27     d=datetime.today()
28     c = ((d.hour * 60) + d.minute) / 20 #what file we are on for the day
29
30     self.filename = "%i-%02d-%02d-%02d.json.gzip"%(d.year,d.month,d.day,c)
31     self.fh = gzip.open(self.basePath + "/" + self.filename,"a")
32     self.tweetCount=0
33     self.errorCount=0
34     self.limitCount=0
35     self.last=datetime.now()
36
37     #This function gets called every time a new tweet is received on the stream
38     def on_data(self, data):
39         datajson = json.loads(data)
40
41         print >> self.fh, json.dumps(datajson)
42         self.tweetCount+=1
43
44         #Check if we should end the stream
45         self.status()
46         return True
47
48     def close(self):
49         try:
50             self.fh.close()
51         except:
52             pass
53
54     def on_error(self, statusCode):
55         print("%s - ERROR with status code %s"%(datetime.now(),statusCode))
56         self.errorCount+=1
57
58     def on_timeout(self):
59         raise TimeoutException()
```

Twitter API

<https://developer.twitter.com/>

3. Profit!

Twitter API

<https://developer.twitter.com/>

3. Profit! Collect data.

```
e_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'ambertrottier', 'id_str': '257519857', 'verified': False, 'followers_count': 343, 'statuses_count': 5318, 'description': 'Proud Booknerd and Book Blogger. Go Bison! 🌿💛 Go Vikings! 💜💜\nIG: @bookobsessed77', 'friends_count': 1921, 'favourites_count': 53927, 'name': 'Amber Trottier 🤘😎', 'lang': 'en', 'location': '', 'time_zone': None}}{'created_at': 'Fri Oct 05 20:29:52 +0000 2018', 'lang': 'en', 'text': 'RT @kaitlancollins: Wow. Heitkamp on her decision to vote no on Kavanaugh: "This isn\'t a political decision. If this were a political decis...', 'truncated': False, 'geo': None, 'place': None, 'favorited': False, 'retweeted': False, 'favorite_count': 0, 'retweet_count': 10053, 'is_quote_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'Kuomon', 'id_str': '245048056', 'verified': False, 'followers_count': 526, 'statuses_count': 16892, 'description': '🎶Wake up, get up, get out there!\nRaise your voice against liars!🎶', 'friends_count': 706, 'favourites_count': 12148, 'name': 'Alfredo Berguido', 'lang': 'en', 'location': 'Panama, Panama', 'time_zone': None}}{'created_at': 'Fri Oct 05 20:29:52 +0000 2018', 'lang': 'en', 'text': 'RT @ABC: Sen. Heidi Heitkamp says she\'ll vote \'no\' on Judge Kavanaugh.\n\nI can\'t get up in the morning and look at the life experience that...', 'truncated': False, 'geo': None, 'place': None, 'favorited': False, 'retweeted': False, 'favorite_count': 0, 'retweet_count': 1026, 'is_quote_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'voter1vote', 'id_str': '374658826', 'verified': False, 'followers_count': 5146, 'statuses_count': 136754, 'description': 'Mom, Grandma, ICU-RN, Fibro Dx ►PLEASE!!, FOR THE LOVE of FREEDOM - VOTE Every Election! ► Prog/Dem, Save our Environment with Science', 'friends_count': 4237, 'favourites_count': 81510, 'name': 'Mrs. Harrington / Call (202) 224-3121', 'lang': 'en', 'location': 'Wisconsin, USA', 'time_zone': None}}{'created_at': 'Fri Oct 05 20:29:52 +0000 2018', 'lang': 'en', 'text': 'RT @Amy_Siskind: The thing that I\'ve taken away in these last 21 months is how few people leading our country, actually care about our coun...', 'truncated': False, 'geo': None, 'place': None, 'favorited': False, 'retweeted': False, 'favorite_count': 0, 'retweet_count': 1363, 'is_quote_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'Hapjamlaw', 'id_str': '24692175', 'verified': False, 'followers_count': 582, 'statuses_count': 166499, 'description': 'RESIST TRUMP! Writer-director; pro-environ., pro-science, pro-indigenous peoples, pro-union, pro-women, pro-LGBTQ / anti-gun, anti-fascist, anti-sexual assault.', 'friends_count': 499, 'favourites_count': 99686, 'name': 'Jim Lawrence', 'lang': 'en', 'location': 'Asheville, North Carolina', 'time_zone': None}}{'created_at': 'Fri Oct 05 20:29:50 +0000 2018', 'lang': 'en', 'text': "RT @brianefallon: Heidi Heitkamp and Lisa Murkowski won't get a fraction of the coverage Flake got for last week's pause, but they showed s...", 'truncated': False, 'geo': None, 'place': None, 'favorited': False, 'retweeted': False, 'favorite_count': 0, 'retweet_count': 2031, 'is_quote_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'brianefallon', 'id_str': '1030000000000000000', 'verified': False, 'followers_count': 1000000, 'statuses_count': 1000000, 'description': 'I am Brian Fallon, a former US Senator from Alaska. I am now a member of the Biden-Harris transition team. I am also a former member of the band The White Stripes. I am also a former member of the band The White Stripes.', 'friends_count': 1000000, 'favourites_count': 1000000, 'name': 'Brian Fallon', 'lang': 'en', 'location': 'Anchorage, Alaska', 'time_zone': None}}
```

Data Cleaning

```
e_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'ambertrottier', 'id_str': '257519857', 'verified': False, 'followers_count': 343, 'statuses_count': 5318, 'description': 'Proud Booknerd and Book Blogger. Go Bison! 🌿💛 Go Vikings! 💜💜\nIG: @bookobsessed77', 'friends_count': 1921, 'favourites_count': 53927, 'name': 'Amber Trottier 🤘😎', 'lang': 'en', 'location': '', 'time_zone': None}}  
{'created_at': 'Fri Oct 05 20:29:52 +0000 2018', 'lang': 'en', 'text': 'RT @kaitlancollins: Wow. Heitkamp on her decision to vote no on Kavanaugh: "This isn\'t a political decision. If this were a political decis...', 'truncated': False, 'geo': None, 'place': None, 'favorited': False, 'retweeted': False, 'favorite_count': 0, 'retweet_count': 10053, 'is_quote_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'Kuomon', 'id_str': '245048056', 'verified': False, 'followers_count': 526, 'statuses_count': 16892, 'description': '♪Wake up, get up, get out there!\nRaise your voice against liars!♪', 'friends_count': 706, 'favourites_count': 12148, 'name': 'Alfredo Berguido', 'lang': 'en', 'location': 'Panama, Panama', 'time_zone': None}}  
{'created_at': 'Fri Oct 05 20:29:52 +0000 2018', 'lang': 'en', 'text': 'RT @ABC: Sen. Heidi Heitkamp says she\'ll vote \'no\' on Judge Kavanaugh.\n\nI can\'t get up in the morning and look at the life experience that...', 'truncated': False, 'geo': None, 'place': None, 'favorited': False, 'retweeted': False, 'favorite_count': 0, 'retweet_count': 1026, 'is_quote_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'voter1vote', 'id_str': '374658826', 'verified': False, 'followers_count': 5146, 'statuses_count': 136754, 'description': 'Mom, Grandma, ICU-RN, Fibro Dx ►PLEASE!!, FOR THE LOVE of FREEDOM - VOTE Every Election! ► Prog/Dem, Save our Environment with Science', 'friends_count': 4237, 'favourites_count': 81510, 'name': 'Mrs. Harrington / Call (202) 224-3121', 'lang': 'en', 'location': 'Wisconsin, USA', 'time_zone': None}}  
{'created_at': 'Fri Oct 05 20:29:52 +0000 2018', 'lang': 'en', 'text': 'RT @Amy_Siskind: The thing that I\'ve taken away in these last 21 months is how few people leading our country, actually care about our coun...', 'truncated': False, 'geo': None, 'place': None, 'favorited': False, 'retweeted': False, 'favorite_count': 0, 'retweet_count': 1363, 'is_quote_status': False, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'Hapjamlaw', 'id_str': '24692175', 'verified': False, 'followers_count': 582, 'statuses_count': 166499, 'description': 'RESIST TRUMP! Writer-director; pro-environ., pro-science, pro-indigenous peoples, pro-union, pro-women, pro-LGBTQ / anti-gun, anti-fascist, anti-sexual assault.', 'friends_count': 499, 'favourites_count': 99686, 'name': 'Jim Lawrence', 'lang': 'en', 'location': 'Asheville, North Carolina', 'time_zone': None}}  
{'created_at': 'Fri Oct 05 20:29:50 +0000 2018', 'lang': 'en', 'text': "RT @brianefallon: Heidi Heitkamp and Lisa Murkowski won't get a fraction of the coverage Flake got for last week's pause, but they showed s...", 'truncated': False, 'geo': None, 'place': None, 'favorited': False, 'retweeted': False, 'favorite_count': 0, 'retweet_count': 2031, 'is_quote_status': True, 'quoted_status_id_str': '1048240930770374656', 'in_reply_to_status_id_str': None, 'in_reply_to_user_id_str': None, 'user': {'screen_name': 'tiger_girl2000', 'id_str': '358279270', 'verified': False, 'followers_count': 1900, 'statuses_count': 49901, 'description': '...feel like I woke up in a parallel universe on November 9, 2016. I just want this country to be a decent place for my kids to grow up in. #TheResistance', 'friends_count': 1803, 'favourites_count': 50625, 'name': 'The Godmother™ 🐾', 'lang': 'en', 'location': 'The Southiest South, USA', 'time_zone': None}}
```

Data Cleaning



Kaitlan Collins

@kaitlancollins

Follow



Wow. Heitkamp on her decision to vote no
on Kavanaugh: "This isn't a political decision.
If this were a political decision for me I
certainly would be deciding this the other
way...I can't get up in the morning and look
at the experience I've had and say yes to
Judge Kavanaugh."

3:13 PM - 4 Oct 2018

10,248 Retweets 47,188 Likes



722



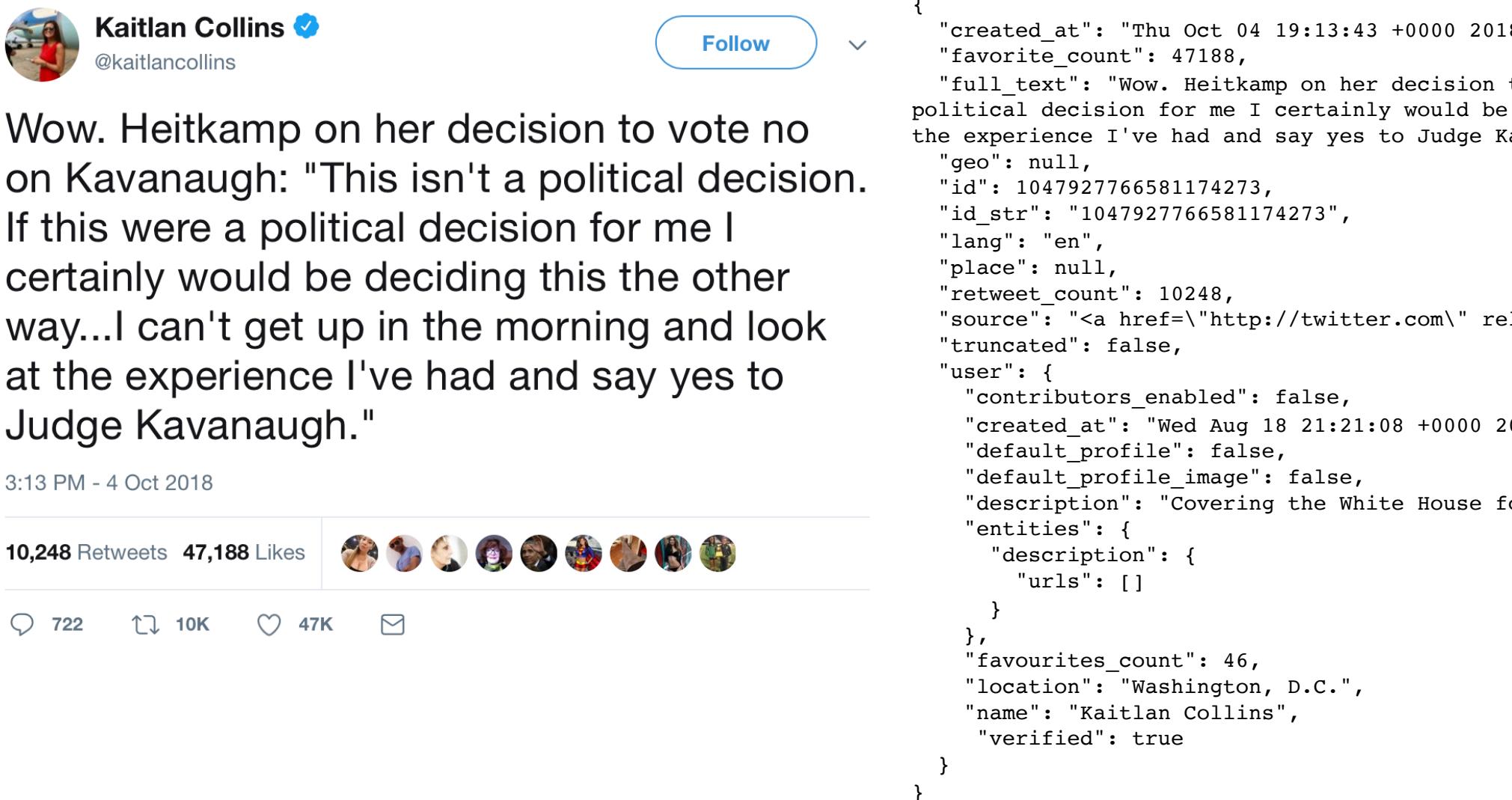
10K



47K



Data Cleaning



Kaitlan Collins 
@kaitlancollins

Follow

Wow. Heitkamp on her decision to vote no on Kavanaugh: "This isn't a political decision. If this were a political decision for me I certainly would be deciding this the other way...I can't get up in the morning and look at the experience I've had and say yes to Judge Kavanaugh."

3:13 PM - 4 Oct 2018

10,248 Retweets 47,188 Likes

722 10K 47K 

```
{
  "created_at": "Thu Oct 04 19:13:43 +0000 2018",
  "favorite_count": 47188,
  "full_text": "Wow. Heitkamp on her decision to vote no on Kavanaugh: \"This isn't a political decision. If this were a political decision for me I certainly would be deciding this the other way...I can't get up in the morning and look at the experience I've had and say yes to Judge Kavanaugh.\n",
  "geo": null,
  "id": 1047927766581174273,
  "id_str": "1047927766581174273",
  "lang": "en",
  "place": null,
  "retweet_count": 10248,
  "source": "<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter for iPhone</a>",
  "truncated": false,
  "user": {
    "contributors_enabled": false,
    "created_at": "Wed Aug 18 21:21:08 +0000 2018",
    "default_profile": false,
    "default_profile_image": false,
    "description": "Covering the White House for @Politico",
    "entities": {
      "description": {
        "urls": []
      }
    },
    "favourites_count": 46,
    "location": "Washington, D.C.",
    "name": "Kaitlan Collins",
    "verified": true
  }
}
```

An Aside: Privacy & Attribution



Kaitlan Collins 

@kaitlancollins

Covering the White House for CNN

📍 Washington, D.C.

🏢 Joined August 2010

An Aside: Privacy & Attribution



Kaitlan Collins 

@kaitlancollins

Covering the White House for CNN

⌚ Washington, D.C.

💻 Joined August 2010



Sarah Shugars

@Shugars

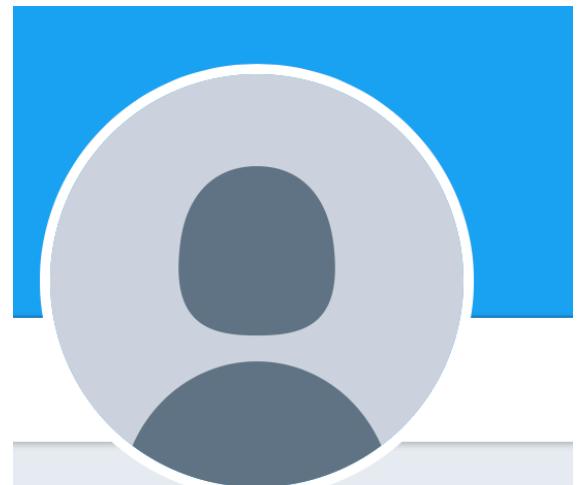
PhD candidate at [@NUNetSI](#).

Computational political scientist studying dialogue & deliberation, socio-semantic networks and natural language processing. She/her.

⌚ Somerville, MA

🔗 sarahshugars.com

💻 Joined February 2009



Overheard at NetSI

@WhileAtNetSI

💻 Joined August 2018

LIFE LESSON 1:

LIFE LESSON 1: TAKE PRIVACY & ATTRIBUTION SERIOUSLY

Okay, Back to Data Cleaning

Wow. Heitkamp on her decision to vote no on Kavanaugh: "This isn't a political decision. If this were a political decision for me I certainly would be deciding this the other way...I can't get up in the morning and look at the experience I've had and say yes to Judge Kavanaugh."

Okay, Back to Data Cleaning

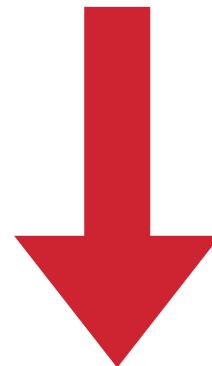
Wow. Heitkamp on her decision to vote no on Kavanaugh: "This isn't a political decision. If this were a political decision for me I certainly would be deciding this the other way...I can't get up in the morning and look at the experience I've had and say yes to Judge Kavanaugh."

- Remove stop words (ie, common words)
- Remove punctuation
- Make lowercase

Okay, Back to Data Cleaning

Wow. Heitkamp on her decision to vote no on Kavanaugh: "This isn't a political decision. If this were a political decision for me I certainly would be deciding this the other way...I can't get up in the morning and look at the experience I've had and say yes to Judge Kavanaugh."

- Remove stop words (ie, common words)
- Remove punctuation
- Make lowercase



wow heitkamp decision vote no kavanaugh political decision
political decision certainly would decide way get morning look
experience say yes judge kavanaugh

Sample Analysis: ‘Heitkamp’ tweets

124,065 Tweets

Collected Oct 4 - Oct 10

Sample Analysis: ‘Heitkamp’ tweets

124,065 Tweets

Collected Oct 4 - Oct 10

Sentiment Analysis

Sample Analysis: ‘Heitkamp’ tweets

124,065 Tweets

Collected Oct 4 - Oct 10

Sentiment Analysis

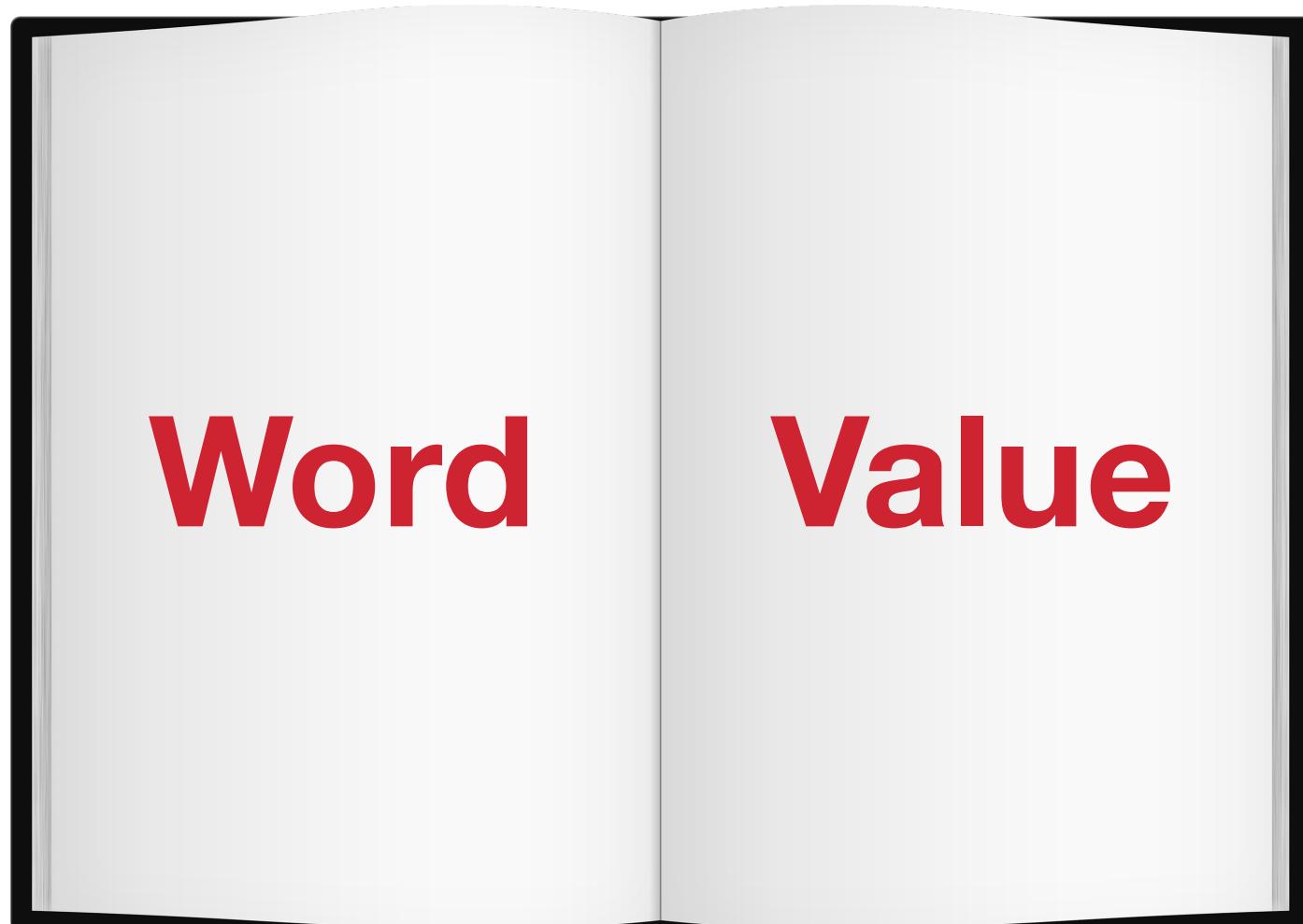


Sentiment Analysis

Dictionary-based:

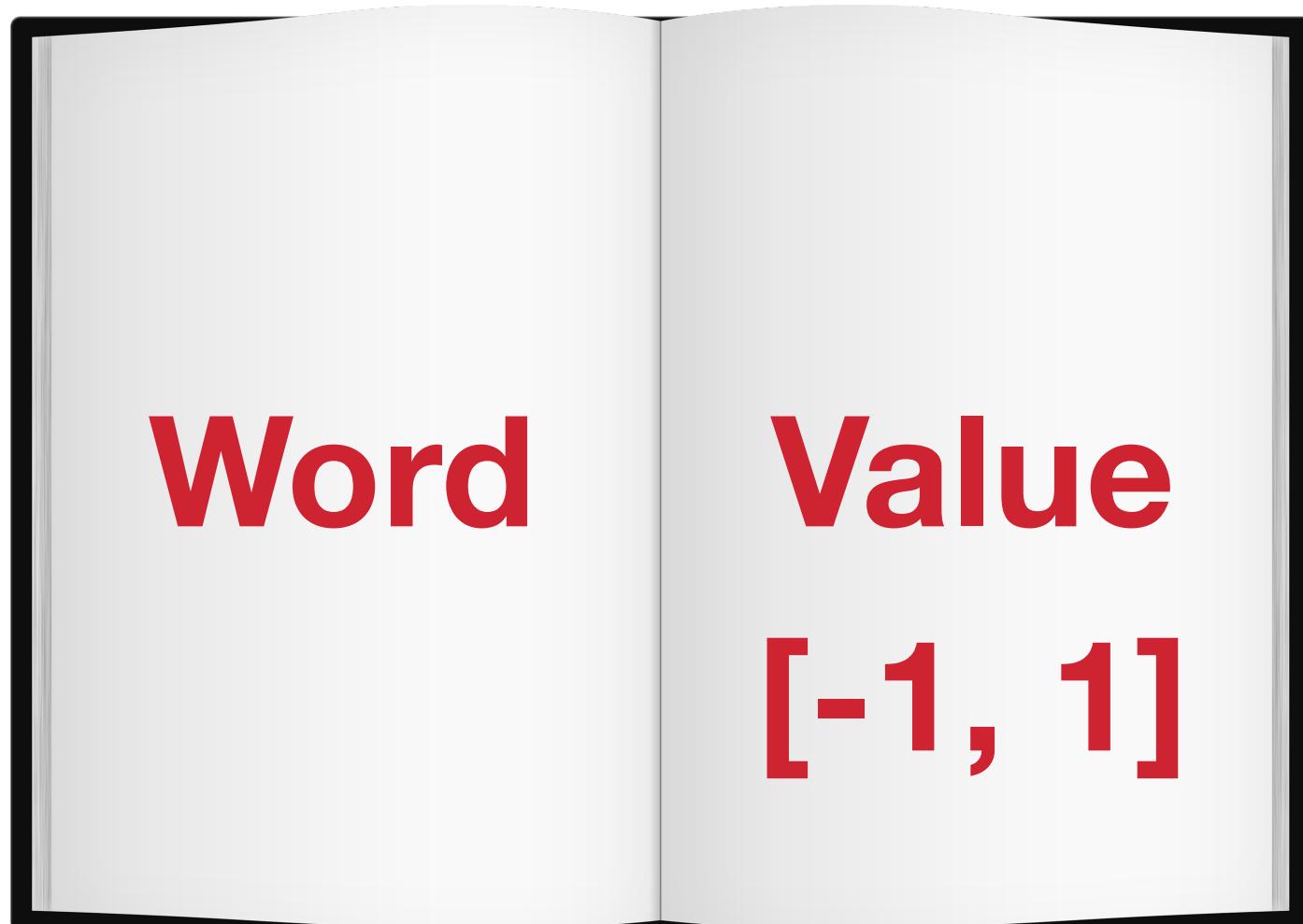
Sentiment Analysis

Dictionary-based:



Sentiment Analysis

Dictionary-based:



Sentiment Analysis

Dictionary-based:

terrible:	good:
-0.4767	0.4404
awful:	great:
-0.4588	0.6249
bad:	awesome:
-0.5423	0.6249

Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

$$\text{Sentiment} = \frac{\text{alexander} + \text{terrible} + \text{horrible} + \text{no} + \text{good} + \text{very} + \text{bad} + \text{day}}{8}$$

Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

$$\text{Sentiment} = \frac{\text{alexander} + \text{terrible} + \text{horrible} + \text{no} + \text{good} + \text{very} + \text{bad} + \text{day}}{8}$$



Bag of Words Approach

Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

$$\text{Sentiment} = \frac{\text{alexander} + \text{terrible} + \text{horrible} + \text{no} + \text{good} + \text{very} + \text{bad} + \text{day}}{8}$$



Bag of Words Approach:
word order doesn't matter

Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

Sentiment =

alexander + terrible + horrible + no + good + very + bad + day

8



Bag of Words Approach:
word order doesn't matter



Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

Sentiment = $\frac{\text{alexander} + \text{terrible} + \text{horrible} + \text{no} + \text{good} + \text{very} + \text{bad} + \text{day}}{8}$

$$= \frac{(0) + (-2.1) + (-2.5) + (-1.2) + (1.9) + (0) + (-2.5) + (0)}{8}$$

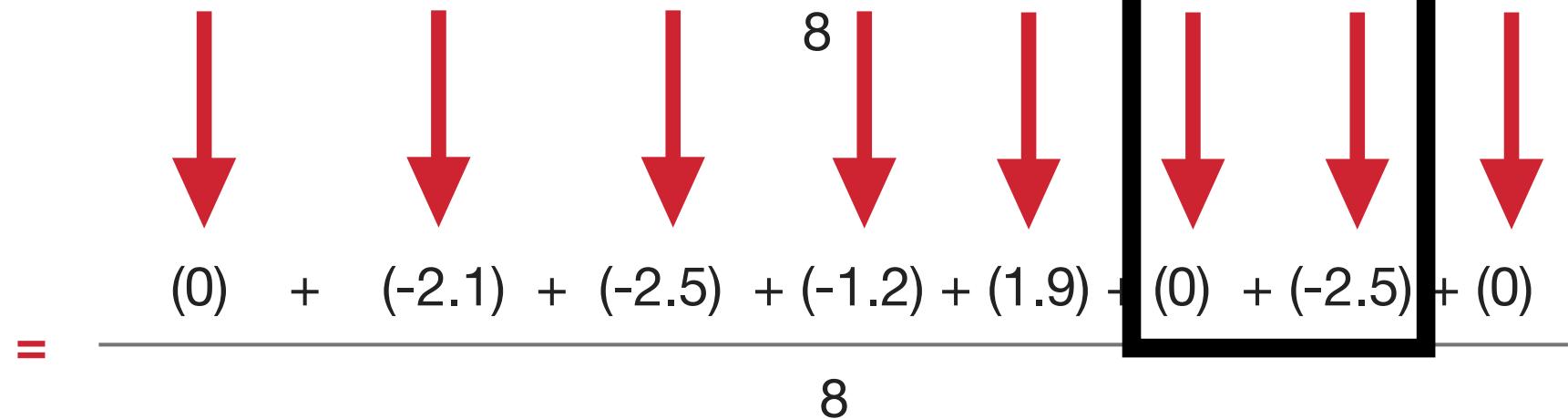
= **-0.8**

Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

Sentiment =

alexander + terrible + horrible + no + good + very + bad + day



Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

Preserve word order

Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

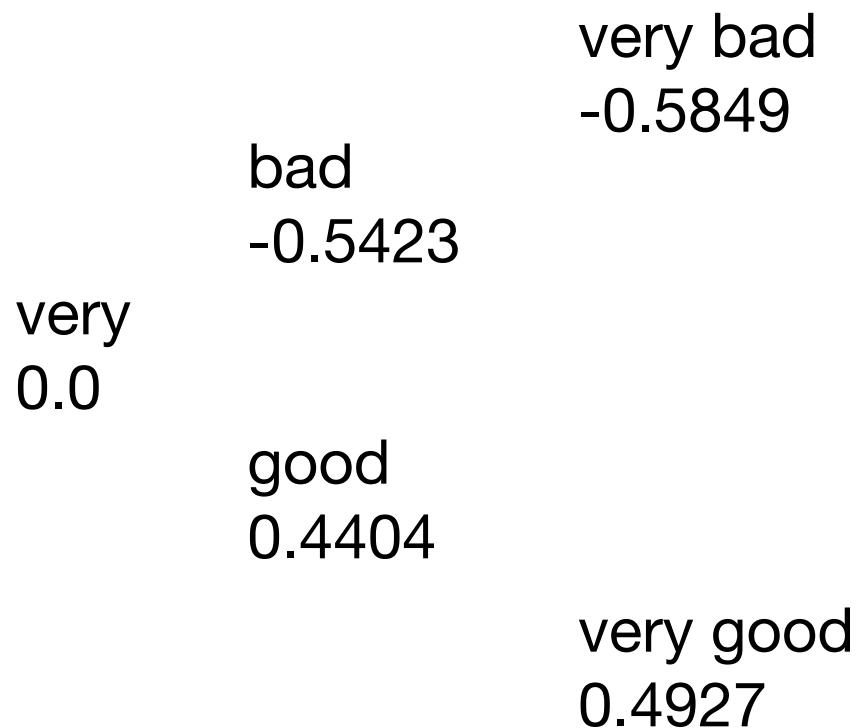
Preserve word order with Vadar



Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

Preserve word order with Vadar

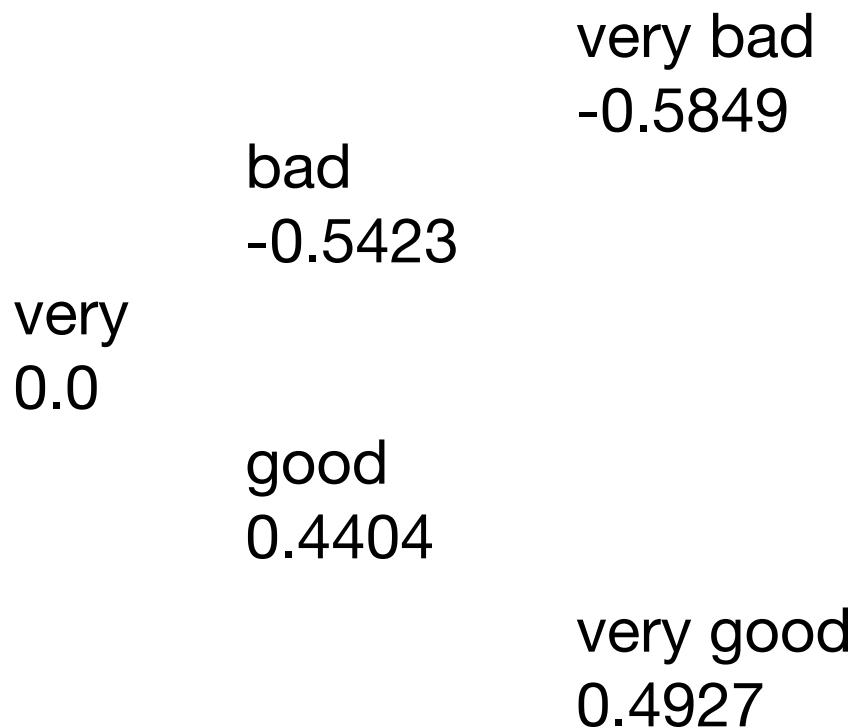


Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

= -0.8 (bag of words)

Preserve word order with Vadar



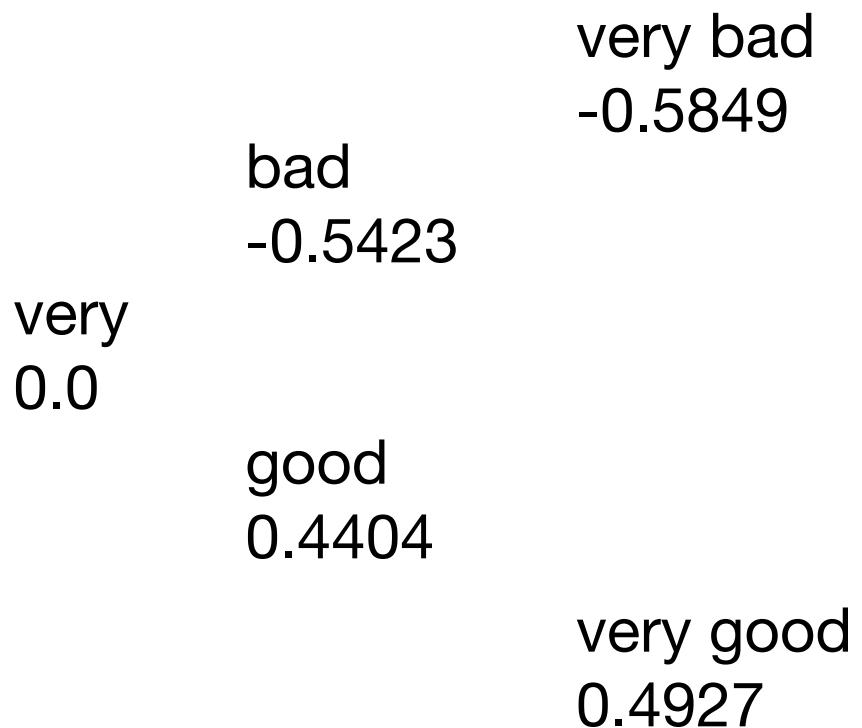
Sentiment Analysis

Alexander and the Terrible, Horrible, No Good, Very Bad Day

= -0.8 (bag of words)

= **-0.8655 (Vadar)**

Preserve word order with Vadar



Sentiment: A Cautionary Tale

Sentiment: A Cautionary Tale



Sentiment: A Cautionary Tale

I didn't sleep at all last night and now I feel just GREAT!

Sentiment: A Cautionary Tale

I didn't sleep at all last night and now I feel just GREAT!

0.729



Sentiment: A Cautionary Tale

I didn't sleep at all last night and now I feel just GREAT!

0.729

The new Tatte is the worst thing that ever happened to me.

Sentiment: A Cautionary Tale

I didn't sleep at all last night and now I feel just GREAT!

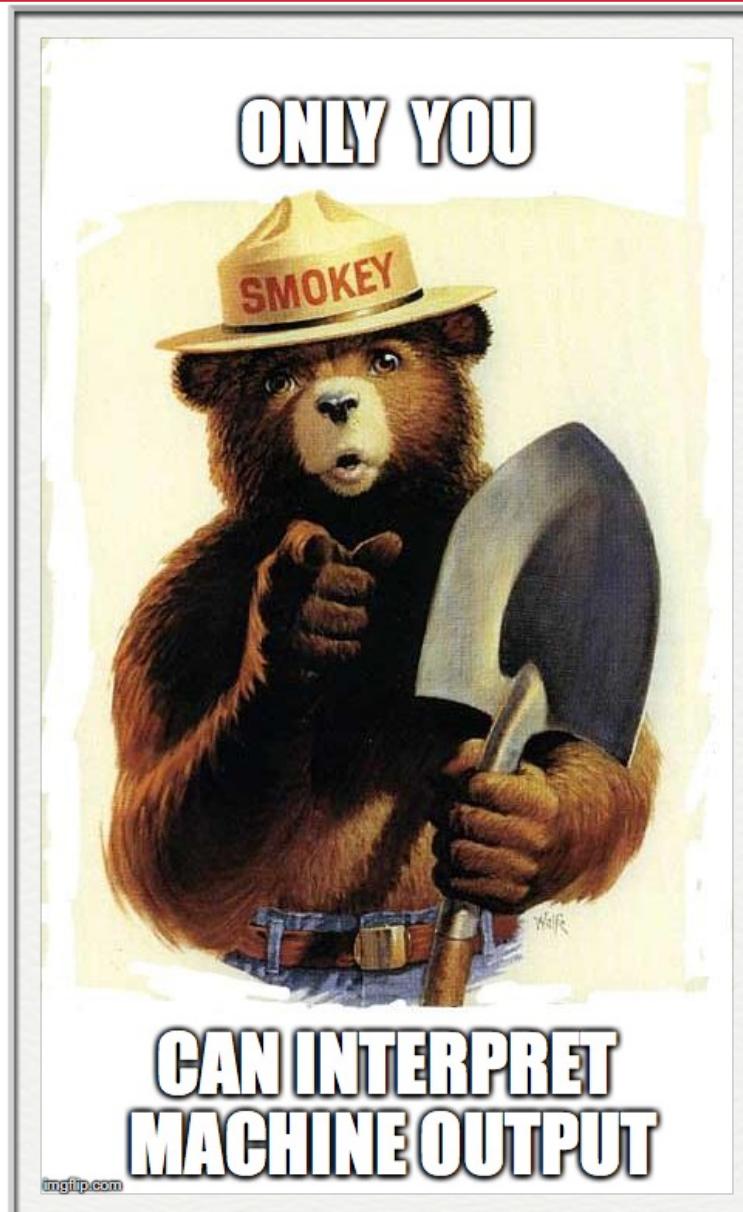
0.729

The new Tatte is the worst thing that ever happened to me.

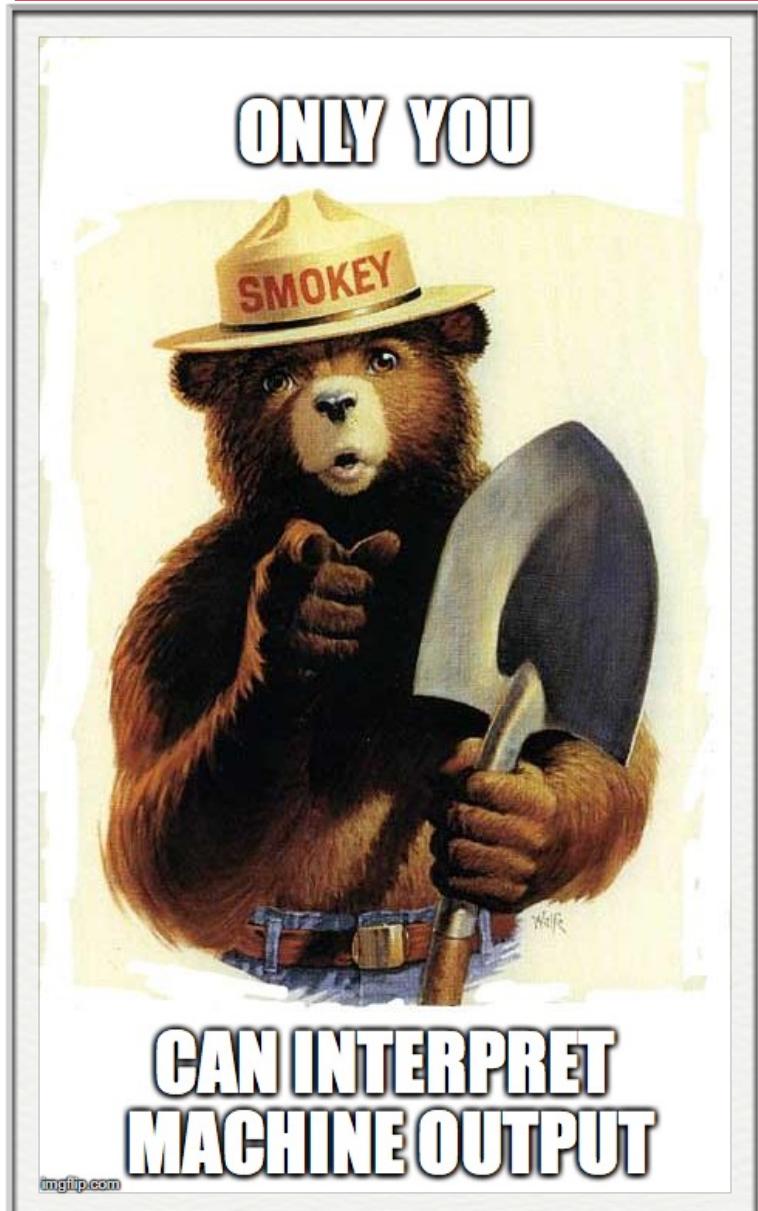
-0.625



Sentiment: A Cautionary Tale

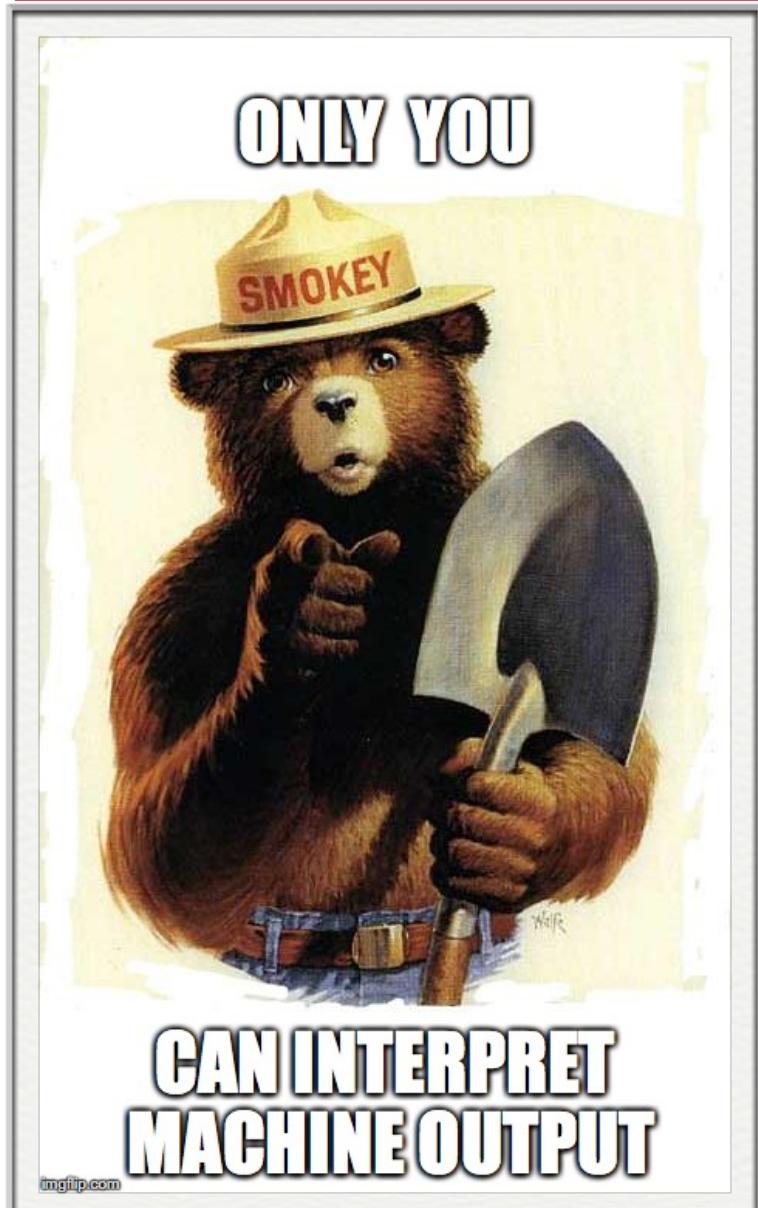


Sentiment: A Cautionary Tale



Algorithms return output

Sentiment: A Cautionary Tale



**Algorithms return output
YOU
have to decide
if the output makes sense**

LIFE LESSON 2:

LIFE LESSON 2:

VALIDATE
MACHINE
OUTPUT

Meanwhile, back at our corpus...

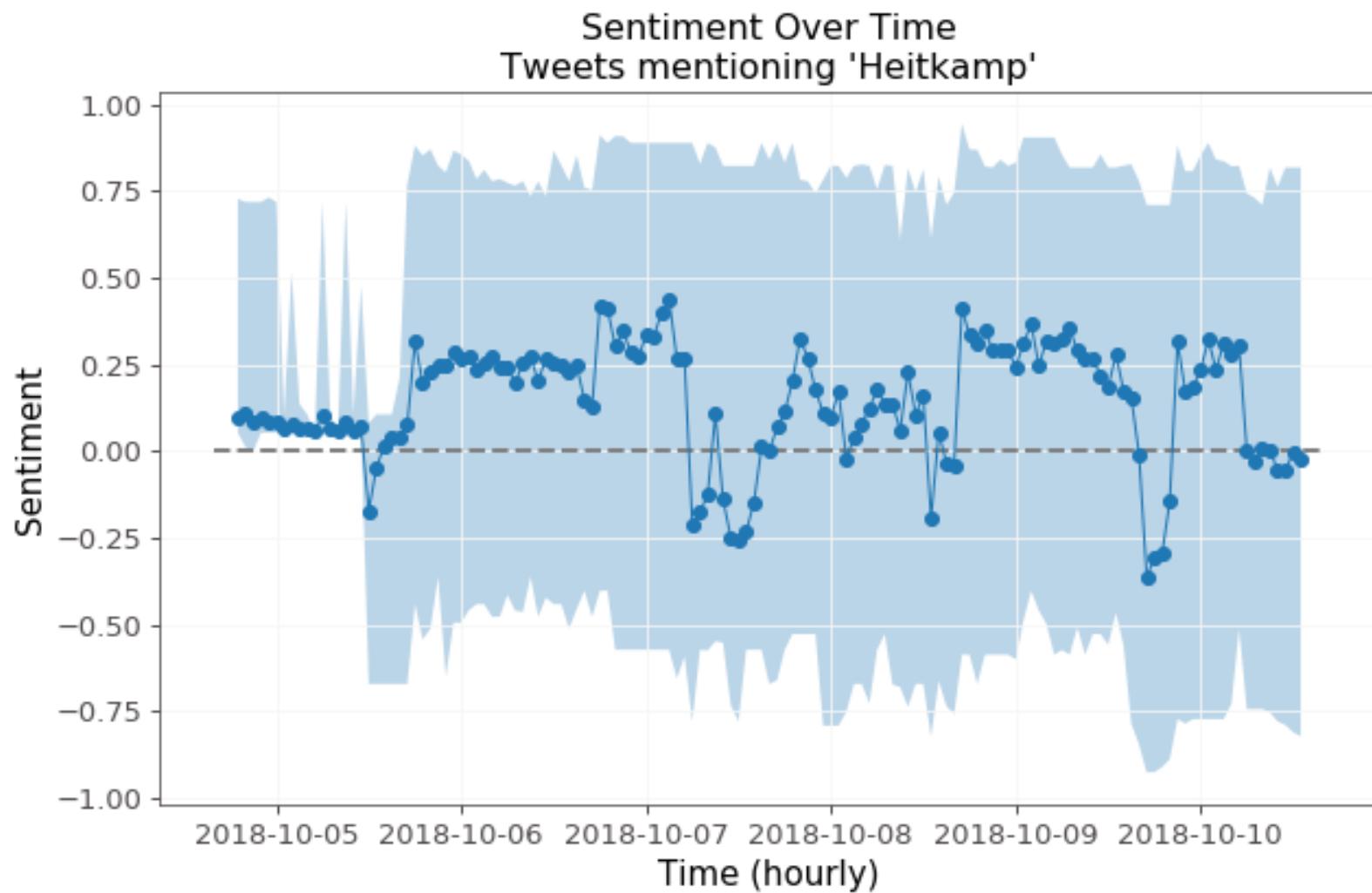
124,065 Tweets

Collected Oct 4 - Oct 10

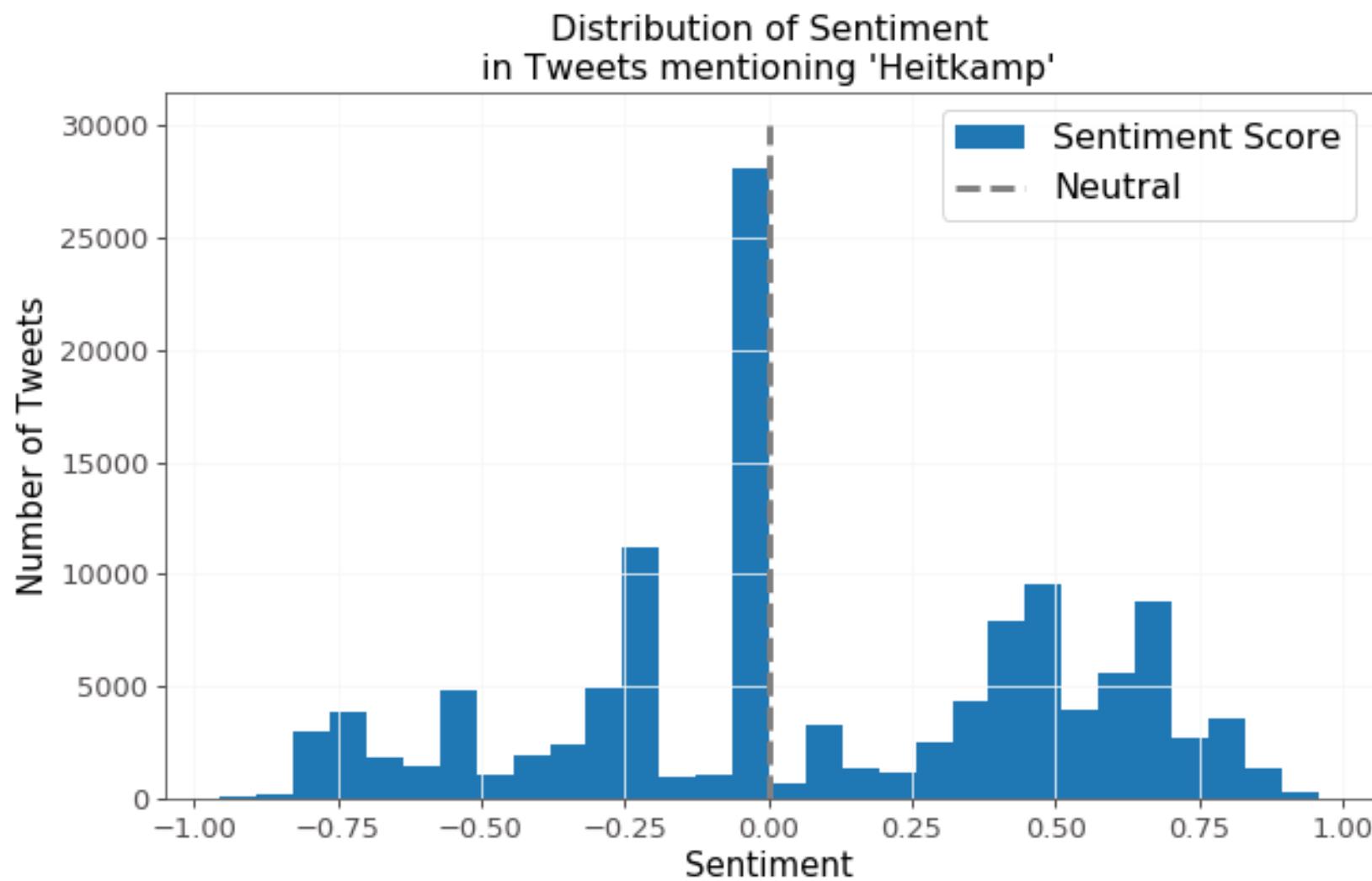
Contain keyword ‘Heitkamp’

Which narrative is dominating?

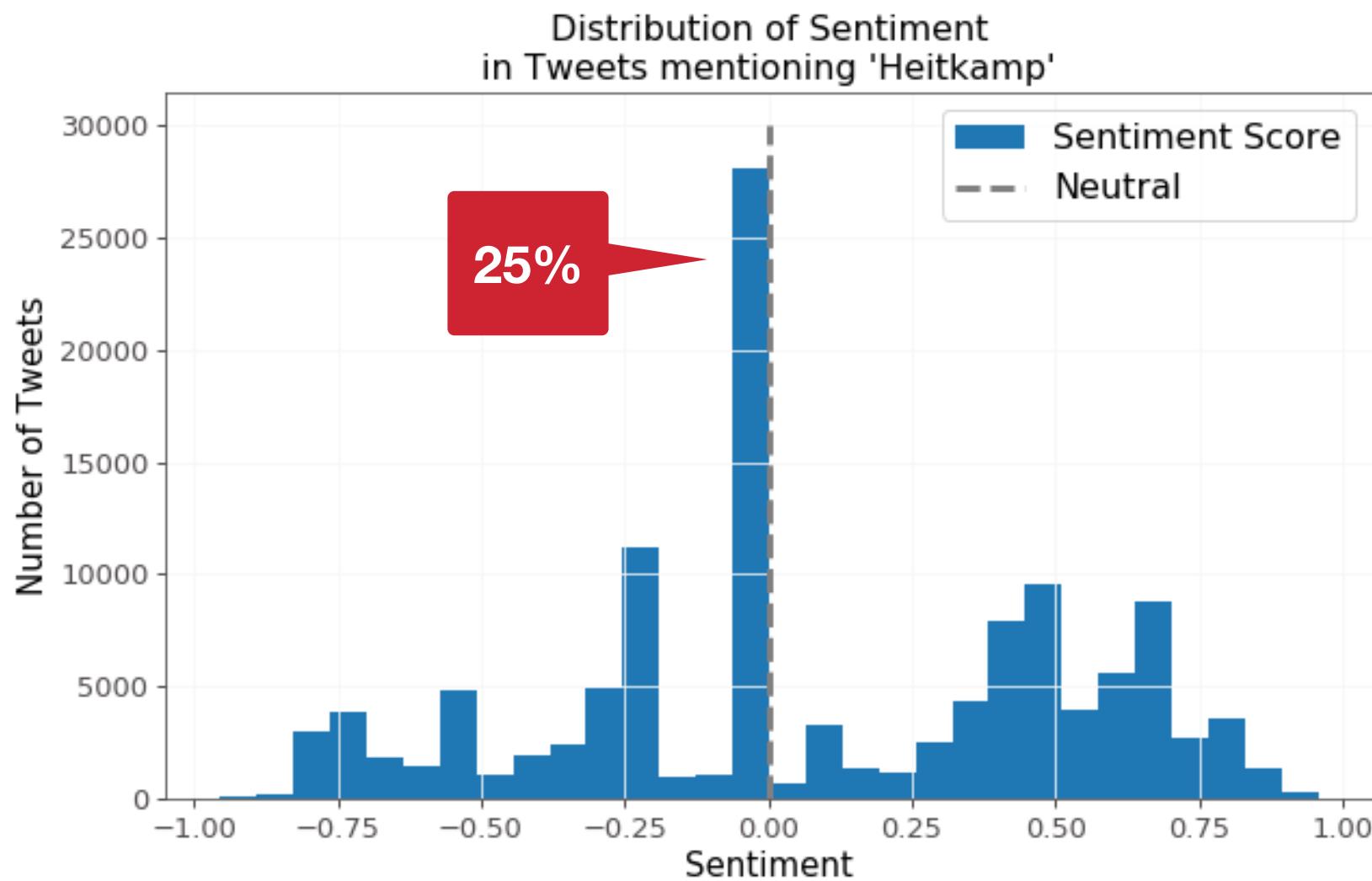
Tweets mentioning 'Heitkamp'



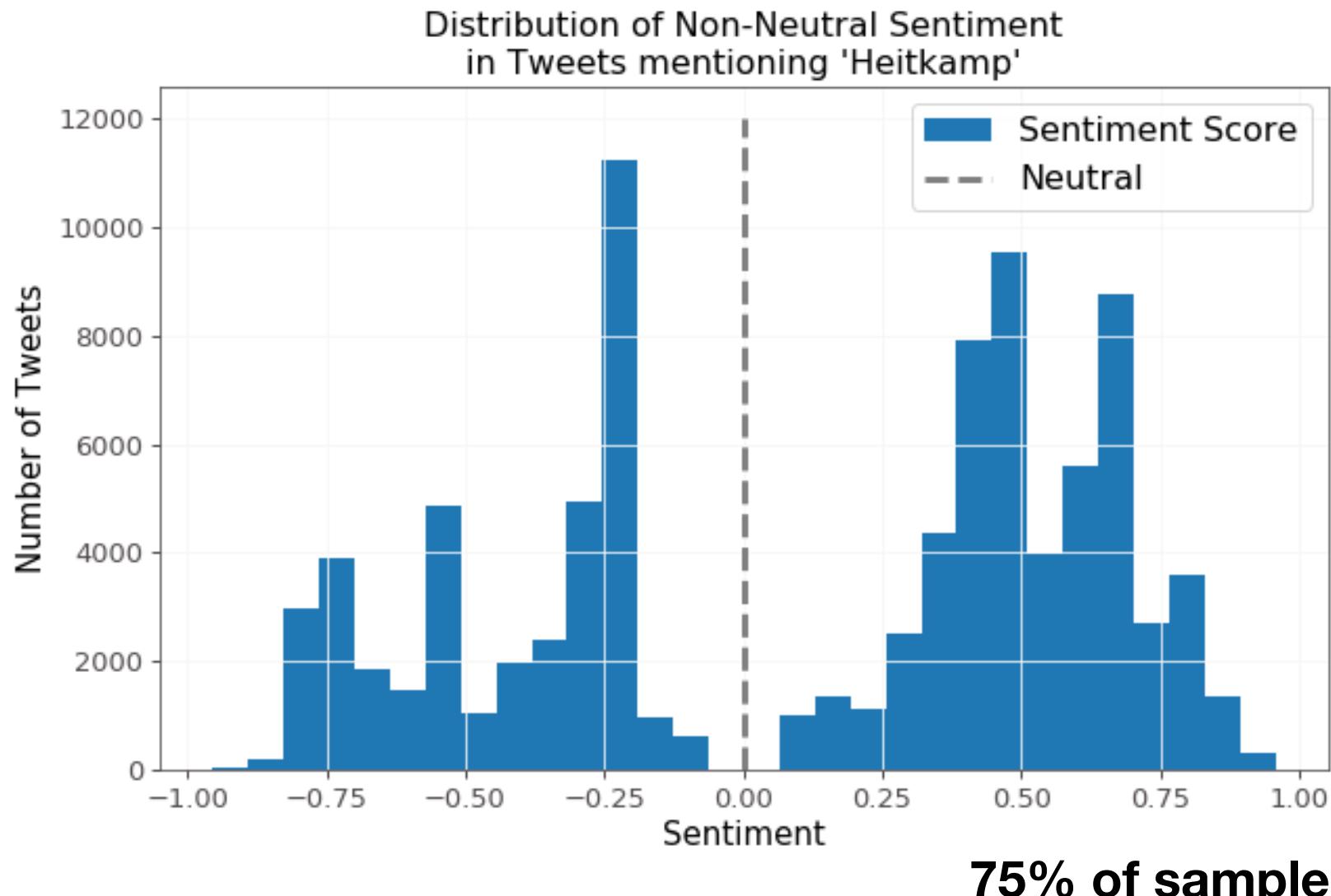
Tweets mentioning 'Heitkamp



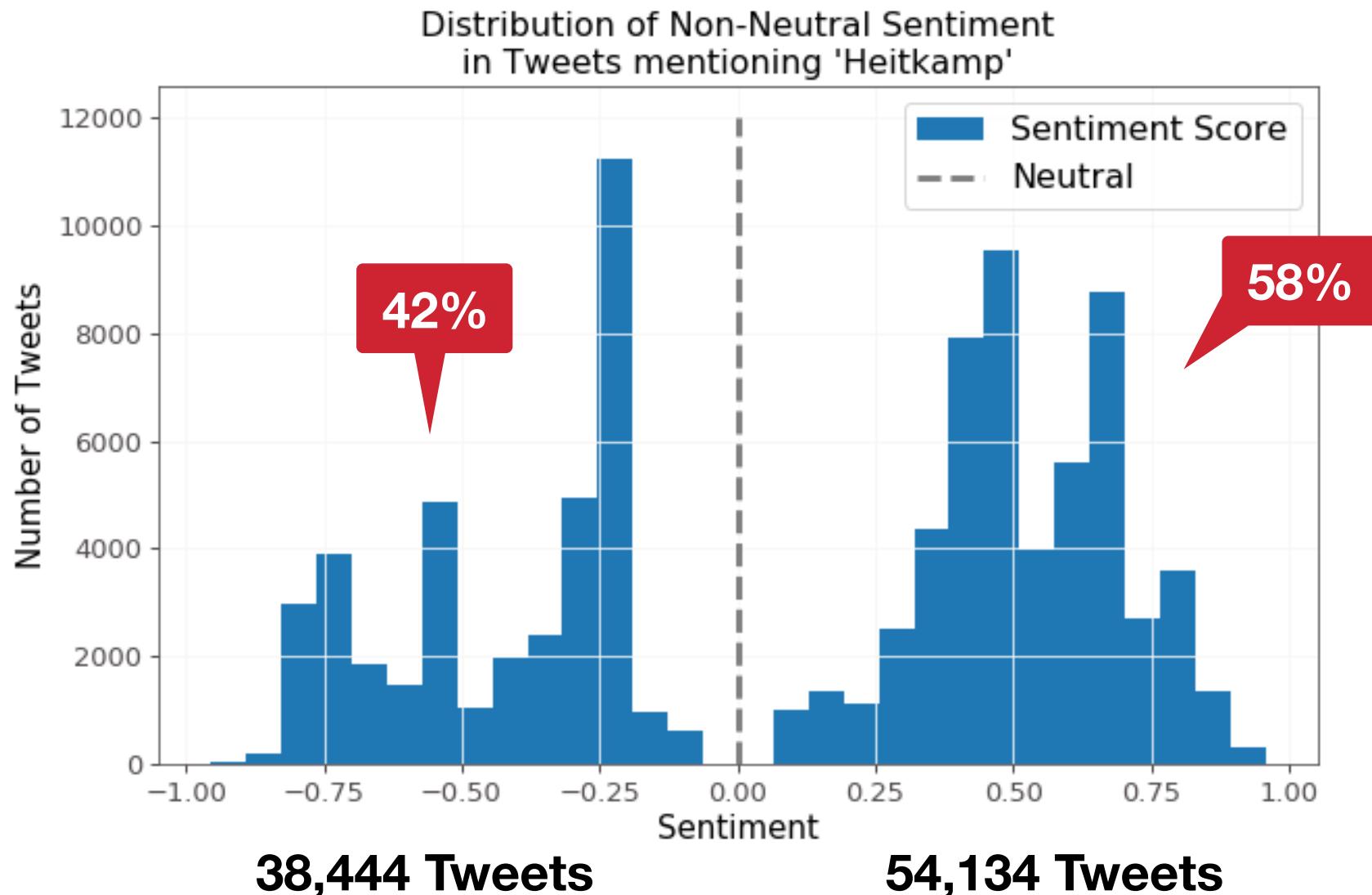
Tweets mentioning 'Heitkamp



Tweets mentioning 'Heitkamp



Tweets mentioning 'Heitkamp



Top Terms

Top Negative Words

- heitkamp (31,279 uses)
- vote (16,748 uses)
- no (14,607 uses)
- heidi (13594 uses)
- kavanaugh (13430 uses)
- senator (10,651 uses)
- judge (9,794 uses)
- decision (9,333 uses)
- nomination (9,245 uses)
- supreme (9,249 uses)

Top Terms

Top Negative Words

heitkamp (31,279 uses)
vote (16,748 uses)
no (14,607 uses)
heid (13594 uses)
kavanaugh (13430 uses)
senator (10,651 uses)
judge (9,794 uses)
decision (9,333 uses)
nomination (9,245 uses)
supreme (9,249 uses)

Top Positive Words

heitkamp (37,765 uses)
vote (26,843 uses)
kavanaugh (16,272 uses)
heid (16,212 uses)
campaign (9,113 uses)
think (7,994 uses)
new (7,983 uses)
honestly (7,922 uses)
told (6,457 uses)
defends (6,442 uses)

Top Terms

Top Negative Words

heitkamp (31,279 uses)

vote (16,748 uses)

no (14,607 uses)

heid (13594 uses)

kavanaugh (13430 uses)

senator (10,651 uses)

judge (9,794 uses)

decision (9,333 uses)

nomination (9,245 uses)

supreme (9,249 uses)

Top Positive Words

heitkamp (37,765 uses)

vote (26,843 uses)

kavanaugh (16,272 uses)

heid (16,212 uses)

campaign (9,113 uses)

think (7,994 uses)

new (7,983 uses)

honestly (7,922 uses)

told (6,457 uses)

defends (6,442 uses)

Top Terms

Top Negative Words

heitkamp (31,279 uses)

vote (16,748 uses)

no (14,607 uses)

heid (13594 uses)

kavanaugh (13430 uses)

senator (10,651 uses)

judge (9,794 uses)

decision (9,333 uses)

nomination (9,245 uses)

supreme (9,249 uses)

Top Positive Words

heitkamp (37,765 uses)

vote (26,843 uses)

kavanaugh (16,272 uses)

heid (16,212 uses)

campaign (9,113 uses)

think (7,994 uses)

new (7,983 uses)

honestly (7,922 uses)

told (6,457 uses)

defends (6,442 uses)

Top Terms

Top Negative Words

heitkamp (31,279 uses)
vote (16,748 uses)
no (14,607 uses)
heidi (13594 uses)
kavanaugh (13430 uses)
senator (10,651 uses)
judge (9,794 uses)
decision (9,333 uses)
nomination (9,245 uses)
supreme (9,249 uses)

Top Positive Words

heitkamp (37,765 uses)
vote (26,843 uses)
kavanaugh (16,272 uses)
heidie (16,212 uses)
campaign (9,113 uses)
think (7,994 uses)
new (7,983 uses)
honestly (7,922 uses)
told (6,457 uses)
defends (6,442 uses)

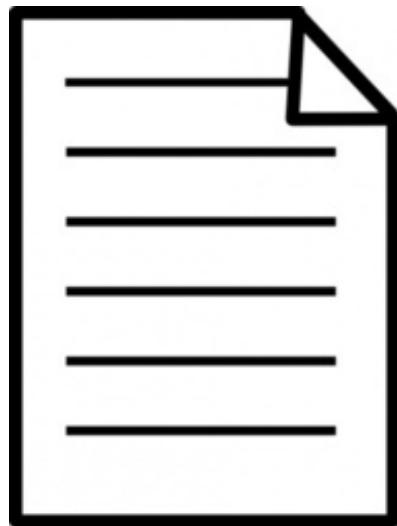
TF-IDF

TF-IDF

Term Frequency-Inverse Document Frequency

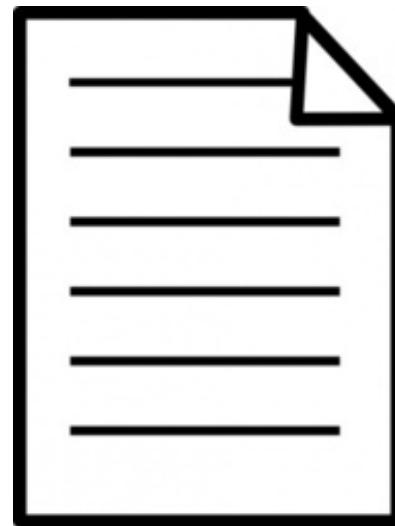
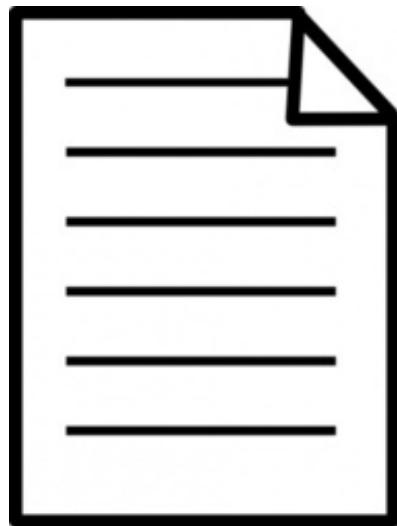
TF-IDF

Term Frequency-Inverse Document Frequency



TF-IDF

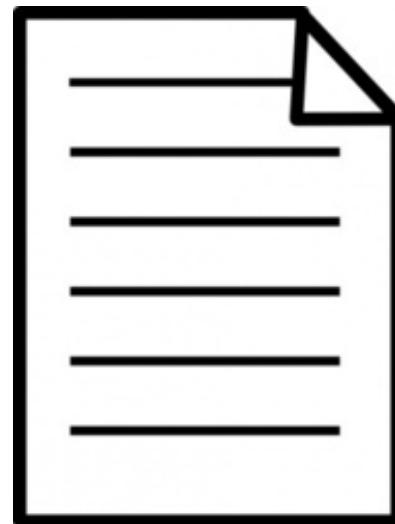
Term Frequency-Inverse Document Frequency



Term t : Appears x time in a document

TF-IDF

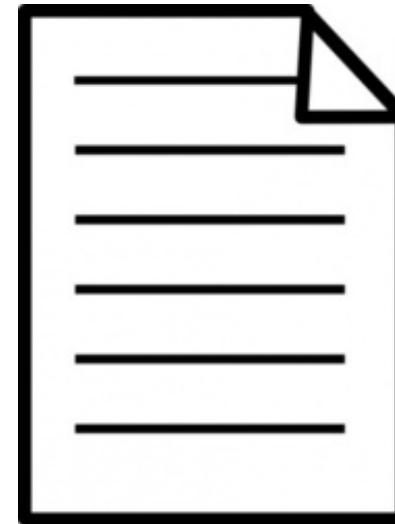
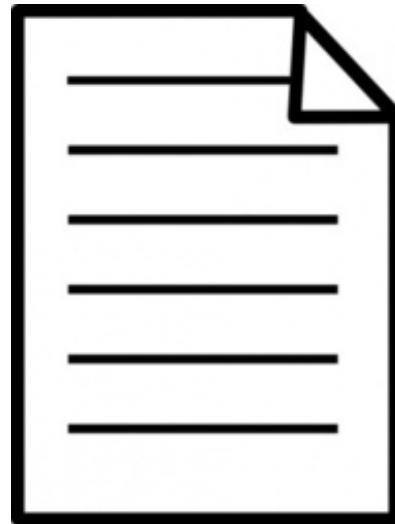
Term Frequency-Inverse Document Frequency



Term t : Appears x time in a document
 Appears y times across documents

TF-IDF

Term Frequency-Inverse Document Frequency



Term t :
$$\frac{\text{Appears } x \text{ time in a document}}{\text{Appears } y \text{ times across documents}}$$

TF-IDF

Term t :
$$\frac{\text{Appears } x \text{ times in a document}}{\text{Appears } y \text{ times across documents}}$$

“heitkamp”:

$$\frac{31,279 \text{ times in negative tweets}}{(31,279 + 37,765) \text{ times in all tweets}} = 0.45$$

TF-IDF

Term t :
$$\frac{\text{Appears } x \text{ times in a document}}{\text{Appears } y \text{ times across documents}}$$

“heitkamp”:

$$\frac{31,279 \text{ times in } \mathbf{\texttt{negative}} \text{ tweets}}{(31,279 + 37,765) \text{ times in all tweets}} = 0.45$$

$$\frac{37,765 \text{ times in } \mathbf{\texttt{positive}} \text{ tweets}}{(31,279 + 37,765) \text{ times in all tweets}} = 0.55$$

TF-IDF

Term t :
$$\frac{\text{Appears } x \text{ time in a document}}{\text{Appears } y \text{ times across documents}}$$

TF-IDF Score

1 : Term only appears in this document

$\frac{1}{\# \text{ of docs}}$: Term appears equally across documents

0 : Term does not appear in this document

Top Terms - Negative Sentiment

Word	TF-IDF	Count
abuse	1.0	4153
unjust	1.0	2664
outrageous	1.0	2658
mainstream	1.0	2327
avenatti	1.0	2327
chuck	1.0	2337
physical	1.0	1699
schoolchildren	1.0	1410
justifiably	1.0	1257
rip	1.0	1163

Top Terms - Negative Sentiment

Word	TF-IDF	Count
abuse	1.0	4153
unjust	1.0	2664
outrageous	1.0	2658
mainstream	1.0	2327
avenatti	1.0	2327
chuck	1.0	2337
physical	1.0	1699
schoolchildren	1.0	1410
justifiably	1.0	1257
rip	1.0	1163



FOXNEWS.COM

Heidi Heitkamp didn't push to prosecute sexual, physical abuse at Native American school in 1990s

North Dakota Sen. Heidi Heitkamp didn't push to prosecute individuals responsible for physical and sexual abuse of students when she was state

Save

Top Terms - Positive Sentiment

Word	TF-IDF	Count
honestly	1.0	7922
defends	1.0	6442
courage	1.0	4266
ready	1.0	2726
strong	1.0	2682
write	1.0	2455
witness	1.0	2364
relatively	1.0	2362
everybody	1.0	2328
win	0.99	3468

Top User Terms

Top terms in bios for users with...

Top User Terms

Top terms in bios for users with...

Negative Sentiment

magaconservative

magalove

molonlabe

λαβε



magamaga

presidenttrump

cruzcrew

makeamericagreatagain

magachristian

Top User Terms

Top terms in bios for users with...

Negative Sentiment

magaconservative
magalove
molonlabe
λαβε

magamaga
presidenttrump
cruzcrew
makeamericagreatagain
magachristian

Positive Sentiment

boycottnra
slightly
glovesoff
folk
nbsp
pagan
followbackresistance
curmudgeon
guncontrol
impeach45

Top User Terms

Top terms in bios for users with...

Negative Sentiment

magaconservative
magalove
molonlabe
λαβε

magamaga
presidenttrump
cruzcrew
makeamericagreatagain
magachristian

Positive Sentiment

boycottnra
slightly
glovesoff
folk
nbsp
pagan
followbackresistance
curmudgeon
guncontrol
impeach45

Neutral Sentiment

scotus
terrible
pronoun
sound
correspondent
gopackgo
intelligence
humanrights
previously
league

LIFE LESSON 3:

LIFE LESSON 3: DATA ANALYSIS IS *ETHICAL* STORYTELLING

LIFE LESSON 3: DATA ANALYSIS IS ***ETHICAL*** STORYTELLING

Summary: Life Lessons

LIFE LESSON 1:

**TAKE PRIVACY &
ATTRIBUTION SERIOUSLY**

LIFE LESSON 2:

VALIDATE  MACHINE  OUTPUT

LIFE LESSON 3:

**DATA ANALYSIS IS
ETHICAL STORYTELLING**

Summary: Tools & Resources

Twitter Scraping

<http://blogtrackers.fulton.asu.edu/> : Search & download tweets

<https://ifttt.com> : Download tweets to a Google sheet

<https://developer.twitter.com/> : DIY with the Twitter API

<https://botometer.iuni.iu.edu/> : Bot detection

Text Analysis

https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/
: Twitter search with sentiment analysis

<http://www.opencalais.com/opencalais-demo/>
: Named entity recognition (proper names, titles, etc)

<http://www.laurenceanthony.net/software/antwordprofiler/>
: Word counts and complexity metrics

Summary: Tools & Resources

These Slides

http://sarahshugars.com/midterm_mayhem.pdf

Code/Data

https://github.com/sshugars/Midterm_Mayhem_code_data

Me

Sarah Shugars

sarahshugars.com

shugars.s@husky.neu.edu