

Making Sense of Twitter

@Bloomberg

DFKI Speech and Language Technology Group Colloquium
November 5, 2020

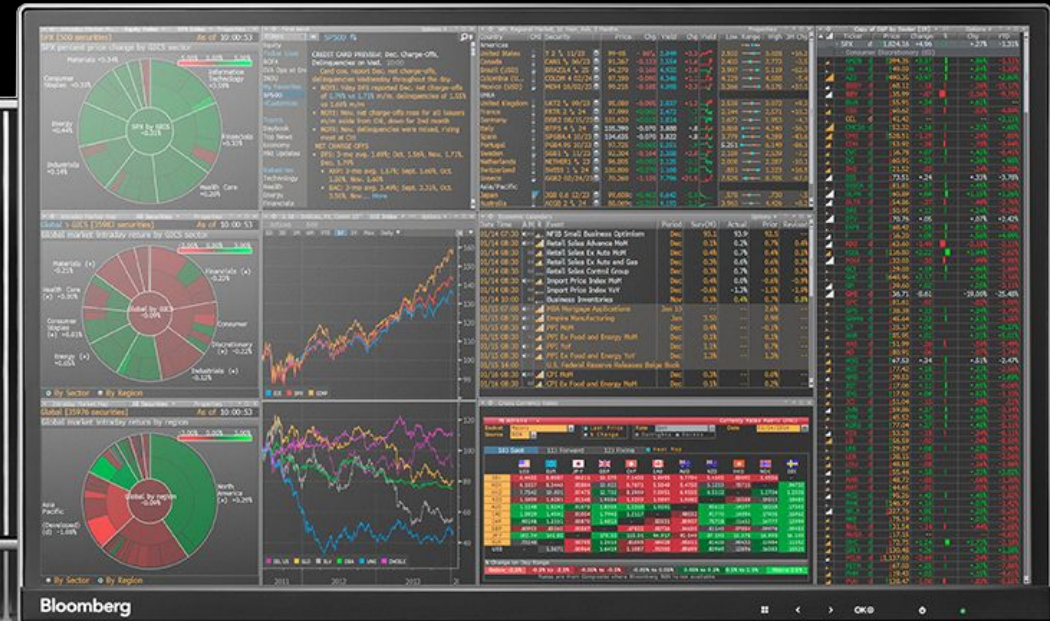
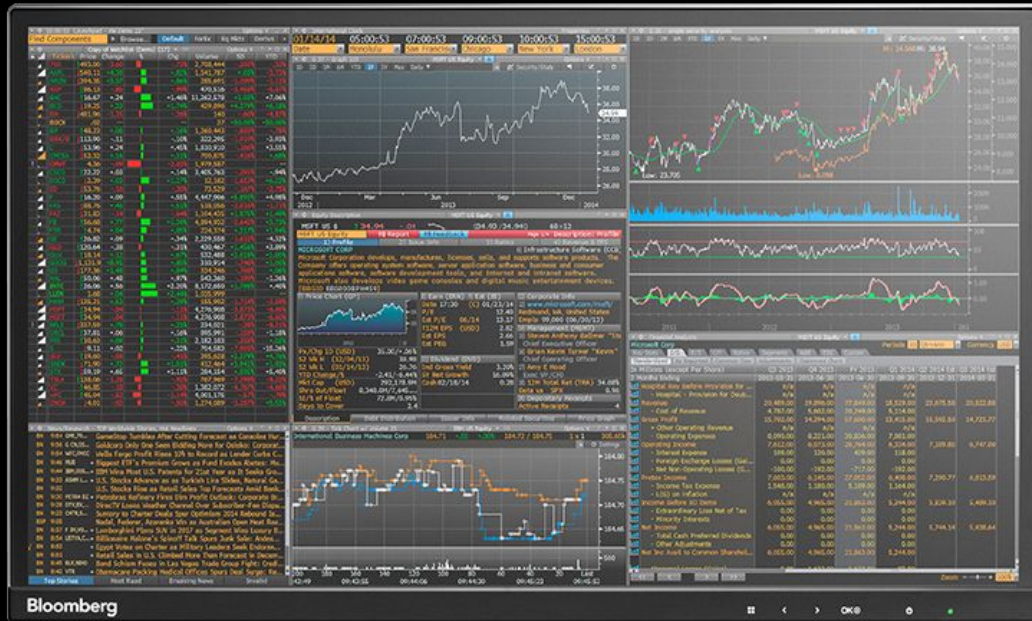
Daniel Preoțiu-Pietro
AI Researcher

TechAtBloomberg.com

© 2020 Bloomberg Finance L.P. All rights reserved.

Engineering

Bloomberg



The **Bloomberg Terminal** is software that delivers a diverse array of information, news and analytics to facilitate financial decision-making.



TechAtBloomberg.com

© 2020 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Bloomberg is just finance, right?

A technology company

Our strength and focus are data, analytics, and community tools

Both creator and consumer of news

~20k employees in 167 offices around the globe

- More than 2,700 journalists and analysts
- 6,000+ software engineers including more than 150 engineers and data scientists working on AI problems

Increased use of and contributions to open source

TechAtBloomberg.com

© 2020 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Why Does Twitter Matter in Finance?

The Bloomberg Terminal is a news platform

Speed

- Quick and easy to share news
- Information spreads quickly

Collective Intelligence

Diversity of sources

- News Articles
- Politicians
- CEOs
- Journalists

SEC approved

- 2013: information significant for company finances can be shared over social media

TechAtBloomberg.com

© 2020 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Source: Politician



Agradezco la oportunidad de haber podido servir a México.

Translated from Spanish by Google

I appreciate the opportunity to have been able to serve Mexico.

Rocked by Resignation Mexican peso tumbles by most since May after finance minister resigns



Source: Bloomberg
USDMXN Curncy (USD-MXN X-RATE) Line Chart 4 Days 4 Minutes

Copyright © 201



-2019 14:51:03

12:15 PM · Jul 9, 2019 · Twitter for iPhone

Source: CEO



John Legere ✓
@JohnLegere

I guess we'll find out ;)

Walter Piecyk ✓ @WaltLightShed · Apr 21, 2015

Verizon's post-paid phone subscriber base dropped 138k in Q1. T-Mobile might be the only one that grows this Q.

10:47 AM · Apr 21, 2015 · Twitter for iPhone

23 Retweets 70 Likes



Apr 21, 2015 11:16:17 [BFW]

Translate To...

T-Mobile CEO Tweets on Phone Subscriber Base Speculation

By Christopher Kingdon

(Bloomberg) -- BTIG analyst [Walt Piecyk](#) says on Twitter "Verizon's post-paid phone subscriber base dropped 138k in Q1. T-Mobile might be the only one that grows this Q."

- T-Mobile CEO [John Legere](#) embeds the analyst's Tweet and says

"I guess we'll find out ;)"

- NOTE: Earlier, VZ said it lost 138k phone subs; reported overall new monthly subs of 565k, boosted by tablets

- NOTE: According to Bloomberg ests., TMUS gained 633.4k postpaid phone subs in qtr, AT&T gained 15.2k, Sprint lost 123k

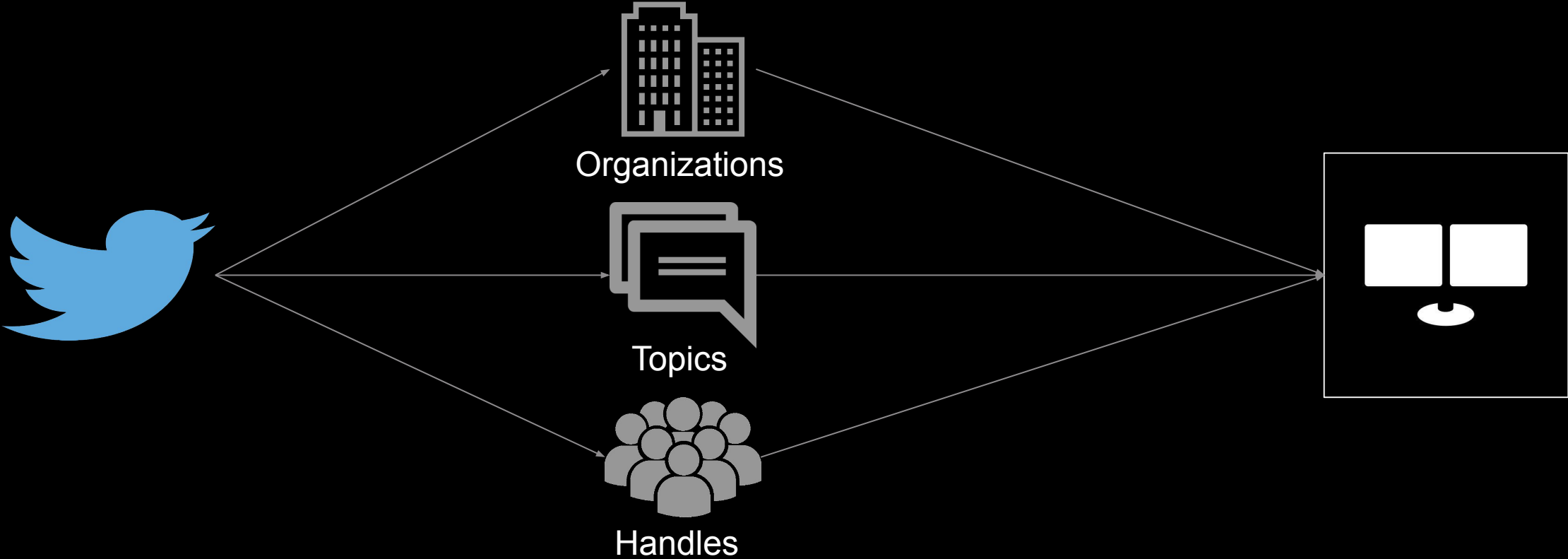
Source: News



The image shows a Twitter post from 'The Associated Press' (@AP). The tweet text reads: 'Breaking: Two Explosions in the White House and Barack Obama is injured'. The post has 3,242 retweets and 153 favorites. It was posted at 12:07 PM on April 23, 2013. The interface includes a 'Following' button and icons for Reply, Retweet, Favorite, Buffer, and More.

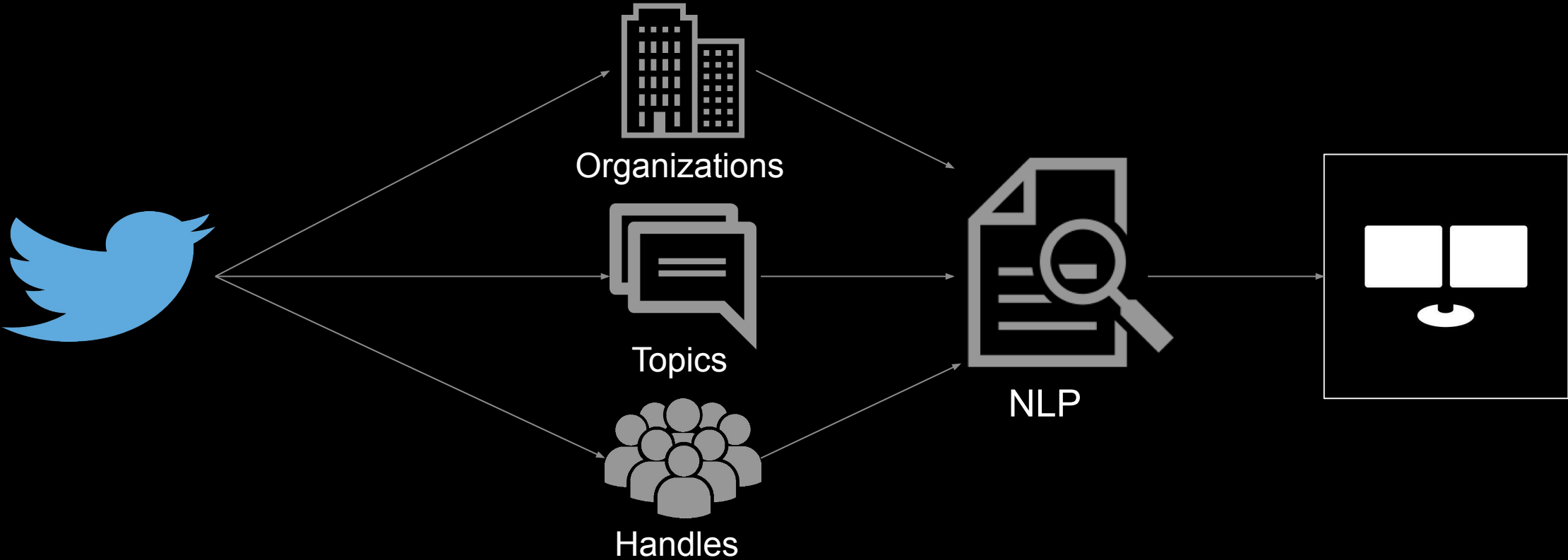
Twitter - Data

Collection is performed around three themes

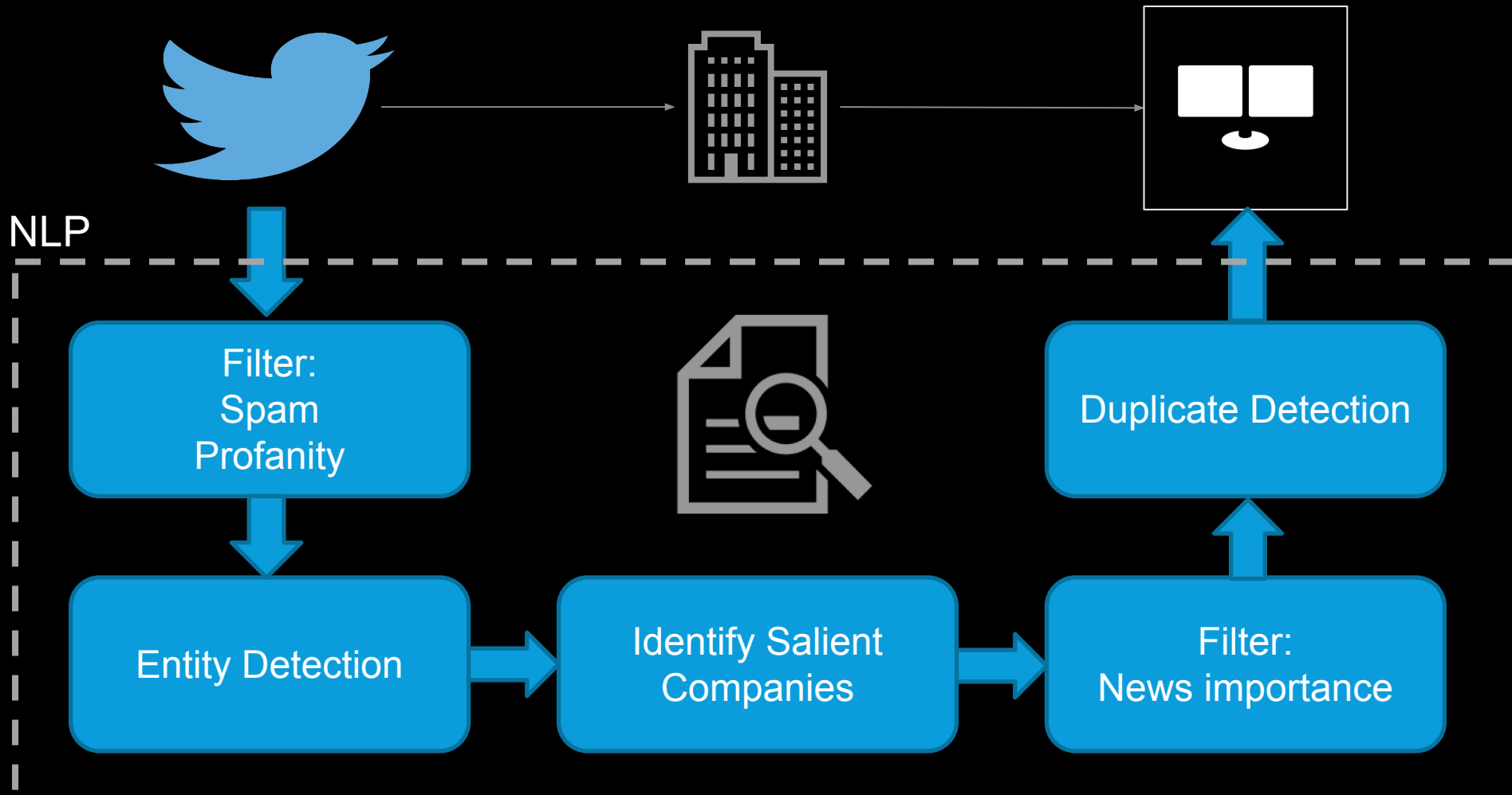


Twitter

NLP is central to all three themes



Twitter – Organizations



Twitter Feed

<Search for a topic, security list, company, or person.> Add Security List

Lists of Twitter Handles Actions

- My Lists
- Featured
 - Bloomberg Suggested
 - U.S. Executive Branch
 - Economics
 - Major News
 - Bloomberg LP
 - Short Focused
- Regions & Countries
- Government & Politics
- Industries
- Markets
- Sports
- Current Events

Search Builder Expected Flow 8.0 Posts/Day


Search Terms

In Addition To within

Lists & Handles


Open in Launchpad

Preview Pop Out

 **NYT Business**
@nytimesbusiness

A \$1 billion investment from Microsoft could help OpenAI, an artificial intelligence research lab in San Francisco, develop technology that does what a human brain can do. But skeptics wonder if it is even possible. nyti.ms/2Y4Wab3


09:08 - 07/22/2019

 **POLITICO**
@politico

China is making serious progress toward its goal of becoming the world leader in artificial intelligence by 2030, raising the question of whether America's hands off approach to technology is enough


politi.co/32zyLSs

10:00 - 07/18/2019

 **NYT Business**
@nytimesbusiness

"There was this incredible optimism," said an analyst at Navigant Research. "Companies thought this was a very straightforward problem. You just throw in some sensors and artificial intelligence, and it would be easy to do."

nyti.ms/2XT3pCQ



Document Clustering and Summarization

News Summaries	
Energy	
Utility Reduce Gas Use After Wednesday Fire	(106 of 52,111 stories)
1) Macomb Daily: Consumers Energy asks customers to reduce natural gas use foll...	NS1 01/31
2) TwinCities.com: Xcel Energy to customers statewide: Turn down thermostat to ...	NS1 01/30
3) Penasee Globe: No heat interruptions yet, but Consumers Energy still asks Mich...	NS1 01/31
DOE Defends Plutonium Shipment to Nevada	(127 of 52,111 stories)
4) Vancouver Sun: DOE secretly shipped plutonium from South Carolina to Nevada	NS2 01/30
5) Tri-City Herald: Judge refuses Nevada's latest bid for plutonium shipment ban	NS1 01/31
6) Brown County Dem: The Latest: Nevada senators blast DOE over plutonium	NS1 01/30
Oil Prices Pushed up by Imports Amid OPEC Efforts	(56 of 52,111 stories)
7) Nigerian Daily: Oil Prices Extend Rise This Month On Signs Of Tighter Supply	NS8 01/31
8) Globe Mail [Reg]: Oil prices fall on doubt over U.S.-China trade talks	GLO 11:51
9) 4-Traders: U.S. crude settles lower amid trade concerns	BLG 01/31
Oil Prices Rise 1 Percent on Venezuela Sanctions	(54 of 52,111 stories)
10) Investing.com: Venezuelan Sanctions Could Mean Oil Prices Rise Sharply In May	FM9 01/31
11) US squeeze on Venezuela oil will reverberate through global energy	FTI 01/31
12) CNBC Wires: FACTBOX-U.S. sanctions on Venezuela's oil industry	NS1 01/30
Samaritan for 70 Chicago Homeless to Stay	(58 of 52,111 stories)
13) stuff.co.nz: Good Samaritan puts up dozens of Chicago's homeless in hotel ami...	NS6 01/31
14) Laredo Times: Good Samaritan picks up hotel tab for 70 homeless in Chicago	NS1 01/31
15) Hartford Courant: The good Samaritans who moved more than 100 people from ...	NS1 03:54
16) Harrisburg P-N: Mystery person picks up hotel tab for 70 homeless during sub-...	NS1 01/31

[NSTM: Real-Time Query-Driven News Overview Composition at Bloomberg](#)

Bambrick, Xu, Almonte, Malioutov, Perarnau, Sello, Chan – ACL 2020

TechAtBloomberg.com

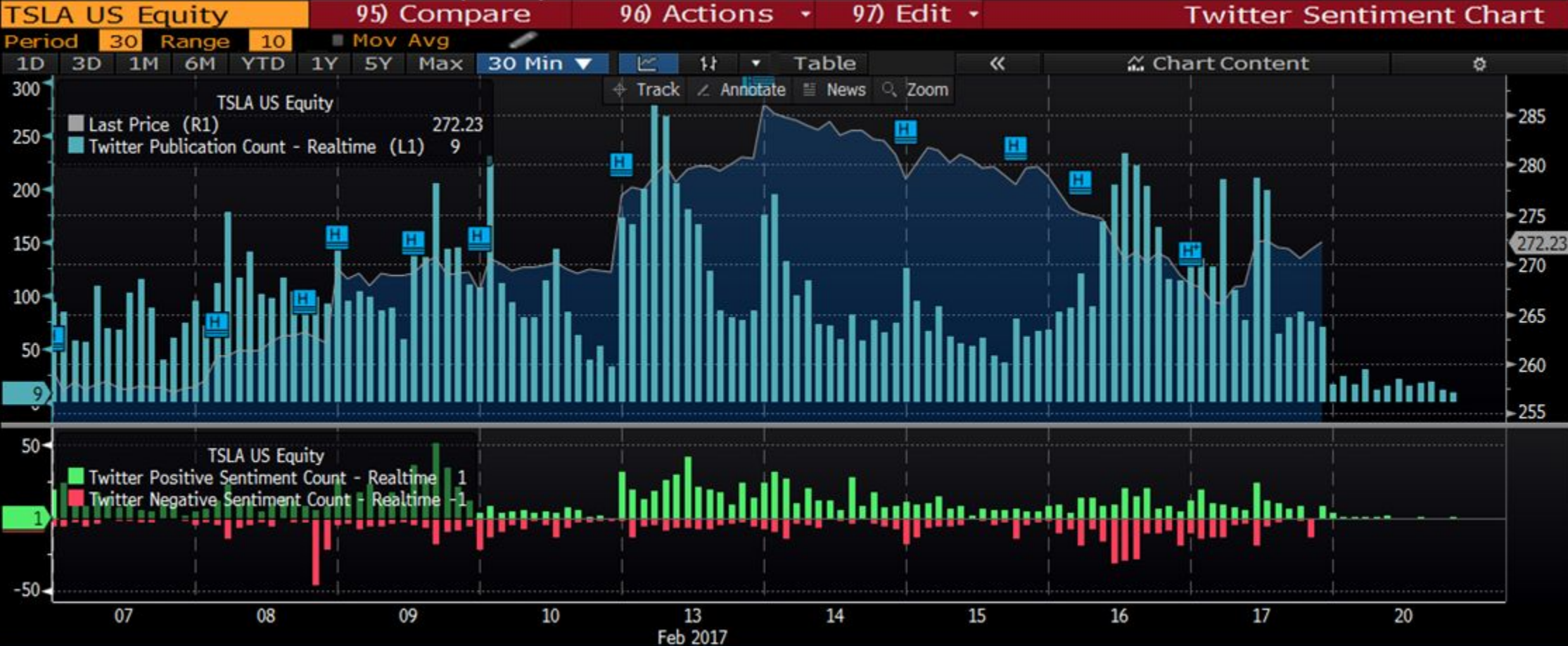
Bloomberg

Engineering

Company Alerts

All Companies		Bloomberg Social Velocity Monitor									
Market Reaction		Stronger									
		Equity				Option			News		
	Time	Company Name	Price	%Chg	Δ AVAT	Volm	Δ Volm	Snmt	Headline		
10)	11:21	GENERAL MOTORS C	30.99	-6.6%	8.7%	13696	-57.5%		Bell Tolls for Car Owners in GM Ignition ...		
11)	09:27	VERINT SYSTEMS	33.49	-4.83%	1891.6%	4176	1073.4%		VERINT STREET WRAP: Shares May Remal...		
12)	09:39	FITBIT INC - A	13.40	-4.35%	80.5%	18184	-19.3%		Fitbit Sinks 7.7%, Continues to Underperf...		
13)	09:35	BOEING CO/THE	128.58	-1.76%	27.1%	30190	50.7%		Boeing's Retooled Plane Unit Aims to Dul...		
14)	12:16	DUKE ENERGY CORP	80.35	-1.1%	-59.9%	707	-82.6%		CharlotteObserver: BRIEF: Duke Energy do...		
15)	10:43	METLIFE INC	44.73	5.35%	247.3%	30755	318.5%		MetLife Gains Capital Flexibility as CEO E...		
16)	10:55	GOLDMAN SACHS GP	156.50	.95%	-24.3%	21428	-.4%		Jefferies Said to Hire Goldman's Ferrenia...		
17)	16:27	MEDIVATION INC	37.39	-3.51%	43.9%	6176	76.9%		Medivation Said to Work With Advisers to ...		
18)	09:46	HIMAX TECHNO-ADR	11.66	-2.83%	57.7%	12134	274.2%		Seeking Alpha: Himax's recent strength a ...		
19)	16:20	PROGRESS SOFTWARE	25.59	1.11%	27.4%	885	1576.1%		Progress 2016 Rev. View Misses Ests., CF...		
20)	10:55	NETFLIX INC	102.19	-1.86%	2.4%	129590	-8.3%		Q. and A.: Ashton Kutcher Talks About His...		
21)	10:23	BANK OF AMERICA	13.48	.45%	-23.9%	210667	-29.0%		Bank of America Top Bond Salesman Brya...		
22)	03:29	AMGEN INC	149.48	-.01%	-41.0%	5749	-46.7%		Fidelity Health Care Adds Alexion, Exits D...		
23)	09:50	ALPHABET INC-C	750.53	.77%	-6.9%	34493	45.9%		Apple Wins Ruling Invalidating One Smart...		
24)	15:39	SQUARE INC - A	15.02	9.32%	62.4%	5725	-8.7%		Square Gains as Much as 10%; Mizuho Call...		
25)	14:29	GALENA BIOPHARMA	1.38	13.11%	168.7%	10419	58.4%		RXi Pharmaceuticals Reports Fourth Quart...		
26)	03:50	TOSHIBA CORP	207.00	-2.13%	-27.8%	--	--		Toshiba Recalls Laptop Computer Battery ...		
27)	09:53	JOHNSON&JOHNSON	108.98	-.15%	-34.2%	14875	-22.5%		Genmab/J&J Multiple Myeloma Study Stop...		
28)	09:23	KELLOGG CO	76.83	.31%	-3.8%	514	-60.8%		Seeking Alpha: Kellogg's: 'They're Grrrr...		
29)	09:48	TWITTER INC	16.36	2.44%	-.2%	84035	4.3%		Turkey's Bozdag Asks Twitter Who Order...		
30)	16:48	MICROSOFT CORP	55.05	.62%	-22.5%	114292	39.6%		Microsoft CEO Stays Committed to AI Bots...		
31)	09:47	FACEBOOK INC-A	114.70	-1.24%	42.1%	363316	54.4%		Facebook Billionaires' Wealth Hits Record...		
32)	15:30	RADIUS HEALTH IN	32.83	6.45%	164.1%	1046	8.0%		Radius Health Climbs After Filing Abalopa...		
33)	16:21	ATYR PHARMA INC	4.86	-3.76%	34.5%	--	--		Atyr Pharma Says Resolaris Preliminary ...		

Twitter Sentiment



Twitter Trends

SPX		Set Alert		Set View as Default		News Trends	
Trend Period		8 Hours					
News Reader Activity		News Sentiment		News Volume		Twitter Volume	
Largest Increase		Largest Total					
Security	Δ Publication	↑ GT	Δ Price	Δ AVAT		Sent. Representative Tweet	
1) Kraft Heinz Co/The			-1.84%	-51.94%		-0.11 Wall Street Journal: The story of Kraft Hei...	
2) Citrix Systems Inc			+0.34%	+23.50%		0.00 \$Stocknewsalerts: \$CTXS New SEC Docume...	
3) Franklin Resources Inc			+1.18%	+1.27%		0.00 Daily Political: Franklin Resources Inc. Ha...	
4) Wells Fargo & Co			+0.28%	+7.09%		-1.00 CNBC Now: BREAKING: Wells Fargo says it ...	
5) Conagra Brands Inc			+1.49%	+31.02%		0.00 stockmarketnow: ConAgra CEO hopes bord...	
6) Bristol-Myers Squibb Co			+0.35%	+176.51%		0.00 Wall Street Journal: Carl Icahn has taken ...	
7) QUALCOMM Inc			+0.51%	-20.79%		+0.01 Engadget: Intel and Qualcomm are steadil...	
8) First Solar Inc			+5.11%	+115.90%		-0.75 Bloomberg: First Solar heads into a restru...	
9) Phillips 66			+0.47%	+27.69%		0.00 Phillips 66: Mayor by day, shift supervisor...	
10) Newmont Mining Corp			+1.16%	+2.61%		-0.19 Estimote: Well done @apeppicali beating 1...	
11) Urban Outfitters Inc			+0.62%	+0.11%		0.00 FOX40 News: Is This Retro or Sad? Urban ...	
12) FirstEnergy Corp			+1.03%	+17.73%		0.00 Andrew Scurria: FirstEnergy Corp. reports ...	
13) CarMax Inc			+1.84%	-30.43%		0.00 Zolmax News: Alyeska Investment Group ...	
14) Yahoo! Inc			+0.89%	-13.82%		-0.39 James Cook: RT @alexweprin: Here's the ...	
15) Microchip Technology Inc			+1.12%	+69.97%		0.00 WKRB News: Microchip Technology Incorpo...	
16) Home Depot Inc/The			+1.41%	+81.56%		+0.04 Benzinga.com: Kevin Kelly's Home Depot ...	
17) CSX Corp			+0.76%	-36.01%		0.00 The NOW Team RE/MAX: CSX announces th...	
18) Mead Johnson Nutrition Co			-0.03%	-24.54%		+0.11 Zolmax News: Alyeska Investment Group ...	
19) Bed Bath & Beyond Inc			+0.78%	+46.00%		0.00 InterCooler: 4,845 Shares in Bed Bath & B...	
20) Travelers Cos Inc/The			-0.33%	-6.68%		0.00 Watchlist News: The Travelers Companies,...	
21) Concho Resources Inc			+0.92%	+64.19%		0.00 Conf Call Tran: Concho Resources Inc. Rep...	
22) Visa Inc			+0.53%	+57.98%		0.00 Ophir Gottlieb: \$V Shh... There's edge in ...	

Social Monitor in Terminal

Allows clients to read Twitter through the Terminal



TechAtBloomberg.com

© 2020 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Categorizing and Inferring the Relationship Between the Text and Image of Twitter Posts

Alakananda Vempala

Daniel Preoțiu-Pietro

Engineering

Bloomberg

TechAtBloomberg.com

Motivation

What's the largest difference in Twitter content in 2010 and 2019?

2010



2019



TechAtBloomberg.com

© 2020 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Motivation

What's the largest difference in Twitter content in 2010 and 2019?

Many more tweets contain images

2010



2019



Motivation

What's the largest difference in Twitter content in 2010 and 2019?

Many more tweets contain images

Approximately 12% tweets are now accompanied by images
> 50M/day

Very little is known about the text-image relationship

2010



2019



Motivation

Text and image in a tweet can be related in several ways:

1. Text is the caption of the image



Motivation

Text and image in a tweet can be related in several ways:

2. Text is a comment on the image



Motivation

Text and image in a tweet can be related in several ways:

3. Image only illustrates part of the tweet text



Applications

Automatically hiding images that don't add additional content to a Twitter post to maximize screen estate



Image can be hidden
Two extra tweets can be displayed

Image has additional content

Applications

- Automatically identify tweets that contain images and their captions
- Data can be used for distant supervision for image classification
- Identify tweets that contain creative descriptions for images

Data – Task Definition – Text Task

Aim: Determine whether there is semantic overlap between the context of the text and the image

Some or all of the content words in the text are represented in the image
(Text is represented)



None of the content words in the text are represented in the image
(Text is not represented)



Data – Task Definition – Image Task

Focuses on the role of the image to the semantics of the tweet

Aim: Identify if the image's content contributes with additional information to the meaning of the tweet beyond the text

Image has additional content that represents the meaning of text & image
(Image adds)

Image does not add additional content that represents the meaning of text & image
(Image does not add)



B

Bloomberg LP @Bloomberg · Dec 22, 2017
Through Bloomberg Startup, our employees help to grow strong global leaders of the future by collaborating closely with our non-profit education partners. bloom.bg/2zfVpzx



Data – Annotation

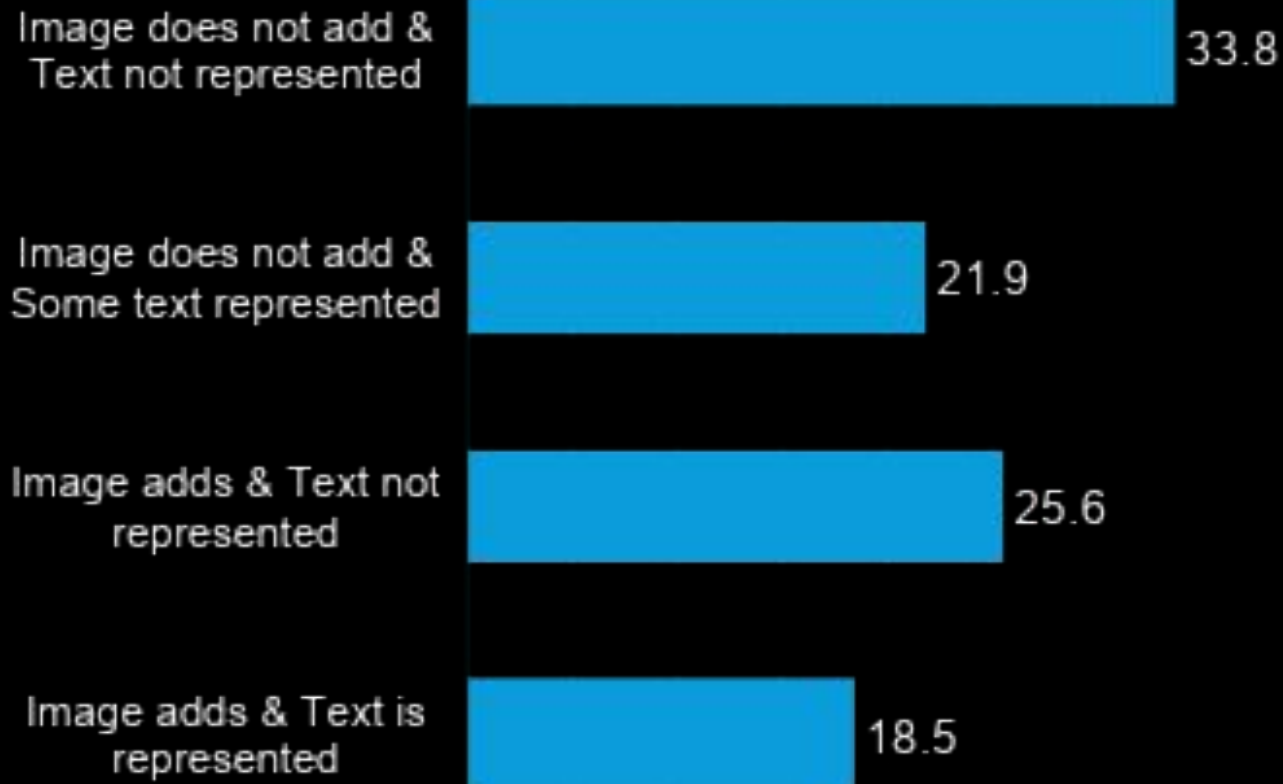
- We used Figure Eight (formerly CrowdFlower)
- Text task
 - Redundancy of 3
 - Krippendorff's Alpha = 0.71 (Text Task)
 - Annotators maintained > 85% accuracy over test questions
- Image task
 - Redundancy of 5
 - Krippendorff's Alpha = 0.46 (Text Task)
 - Annotators maintained > 75% accuracy over test questions
- Adjudication by majority vote

Data – Collection

- Collected annotations for 4,888 tweets
 - All tweets posted in the same year (2016)
 - Split across original posts, retweets and favorited posts
 - Deliberately sampled from handles with known demographic traits
 - Tweets are all in English
- Available online: <https://github.com/danielpreotiuc/text-image-relationship>

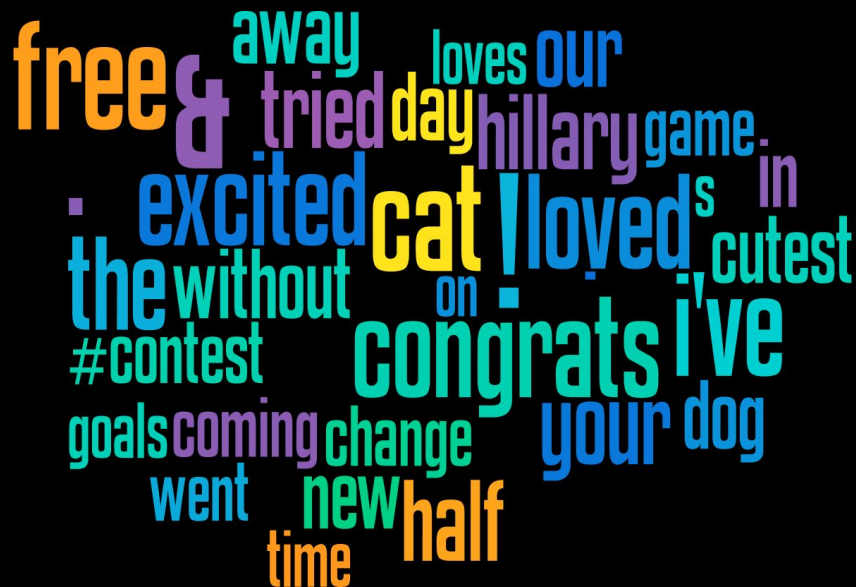
Data – Distribution

Both task labels are combined to assign one of four classes to each text-image pair

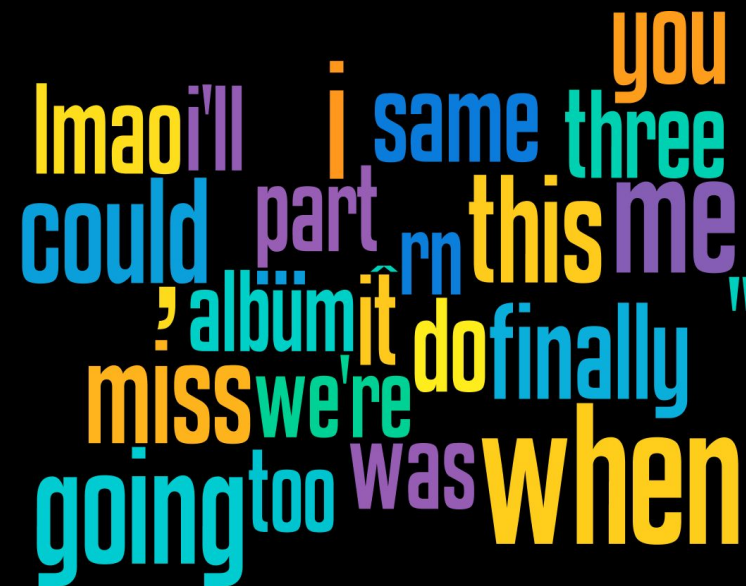


Analysis – Text Task

- Univariate Point-Biserial Correlation between unigram features and text task outcome
- Age is correlated with text being represented in image
 - Especially when image also adds information
 - More *traditional* type of relationship
- Simple tweet metadata not correlated



Some or all of the content words in the text are represented in the image
(Text is represented)



None of the content words in the text are represented in the image
(Text is not represented)

Prediction – Methods

- Demographics
- Metadata
- Text-Based Methods
 - Surface
 - Unigrams
 - BiLSTM
- Image-Based Methods
 - ImageNet Classes
 - InceptionNet Tuned
- Joint Text+Image approaches
 - Linear ensemble of text and image predictions
 - LSTM + InceptionNet architecture

TechAtBloomberg.com

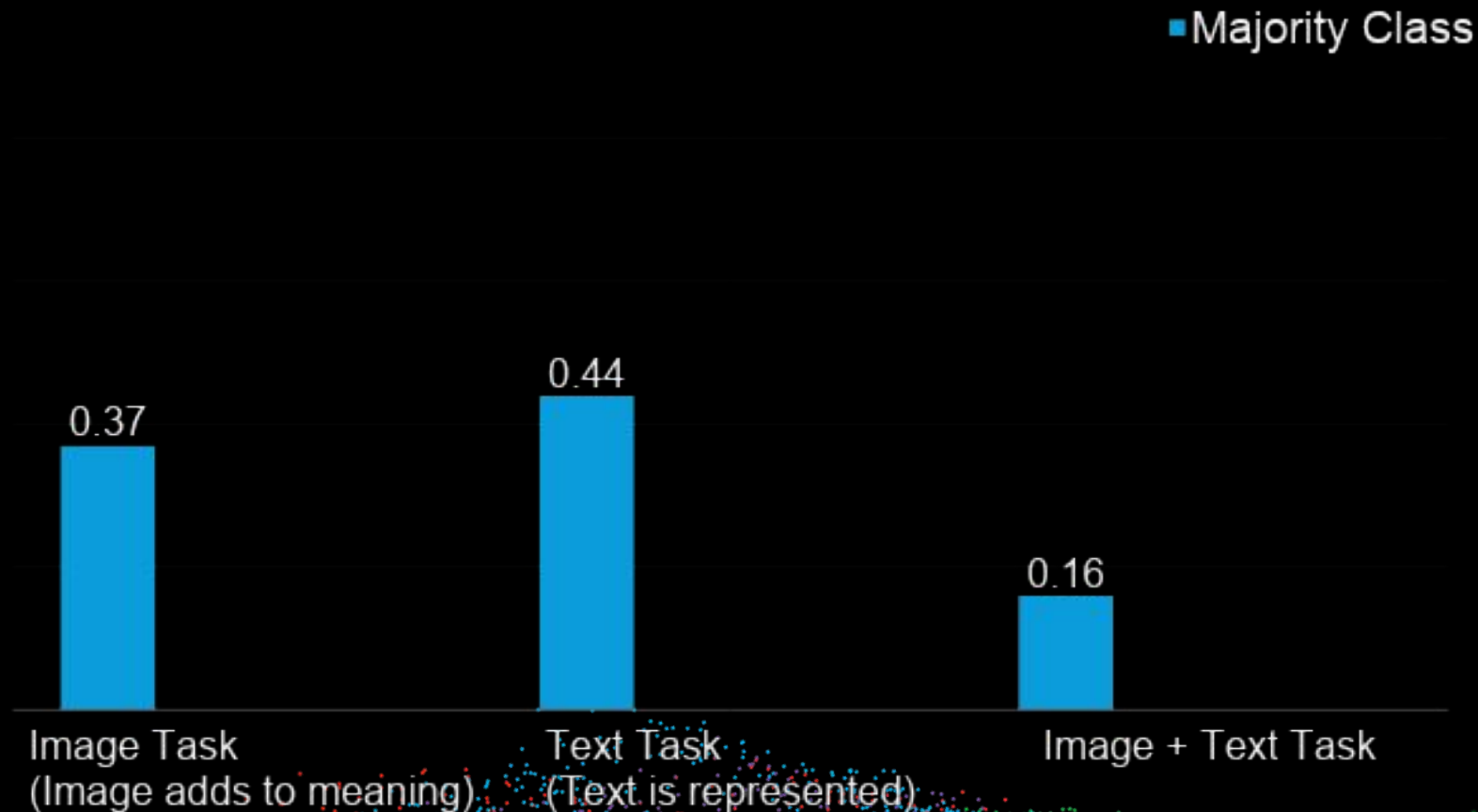
© 2020 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Prediction – Baseline Methods

- Results in weighted F1, Train (80%), Test (20%), Parameters set via 10-fold CV
- Three tasks: Image Task (binary), Text Task (binary), Image + Text (4-class)
- Demographics, tweet metadata features almost no predictive value

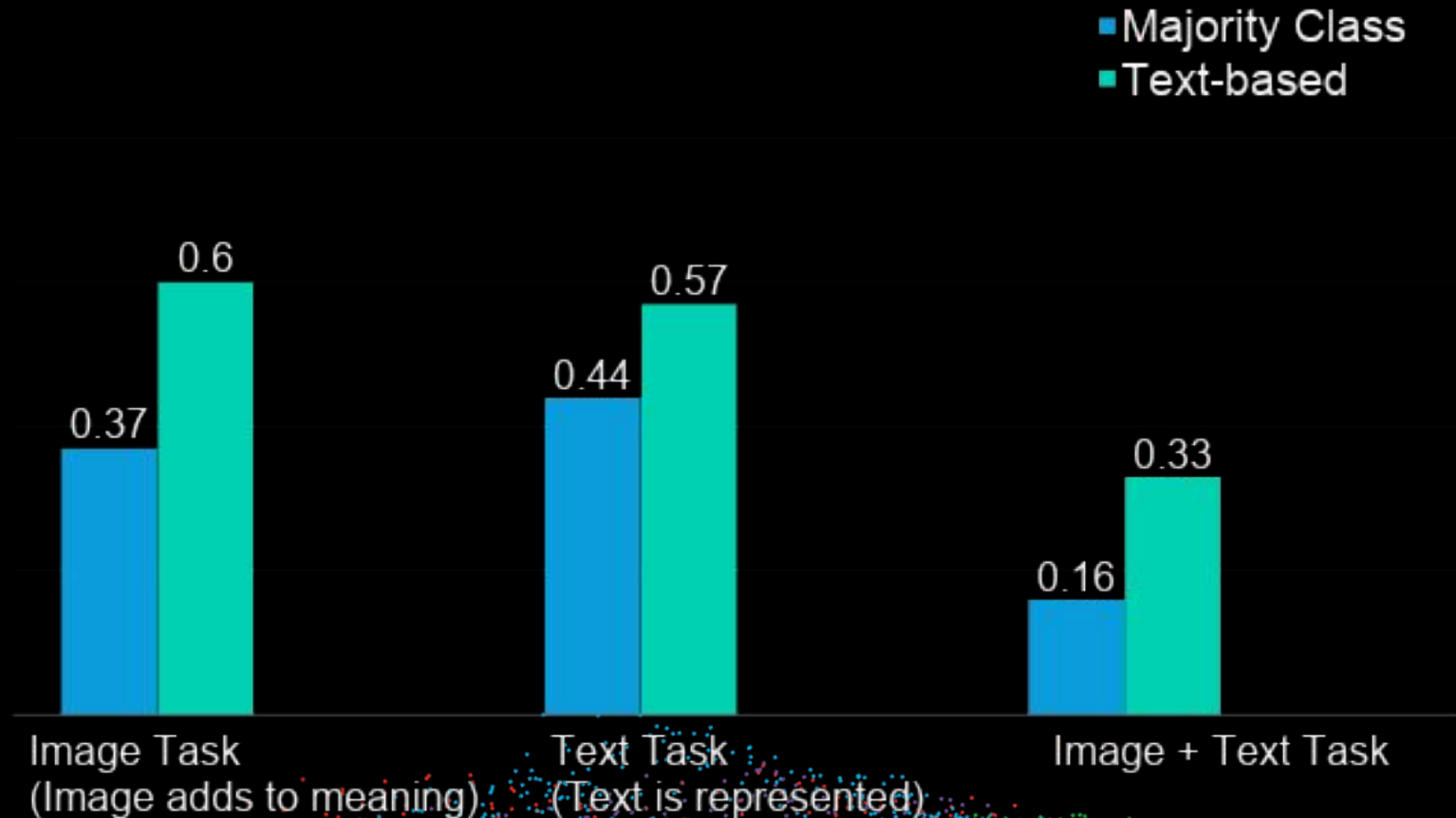


Bloomberg

Engineering

Prediction – Text-based Methods

- All tasks show improvements on the baseline
 - Higher predictive power on image task (does the image have additional content)
- LSTM models are marginally better than unigrams

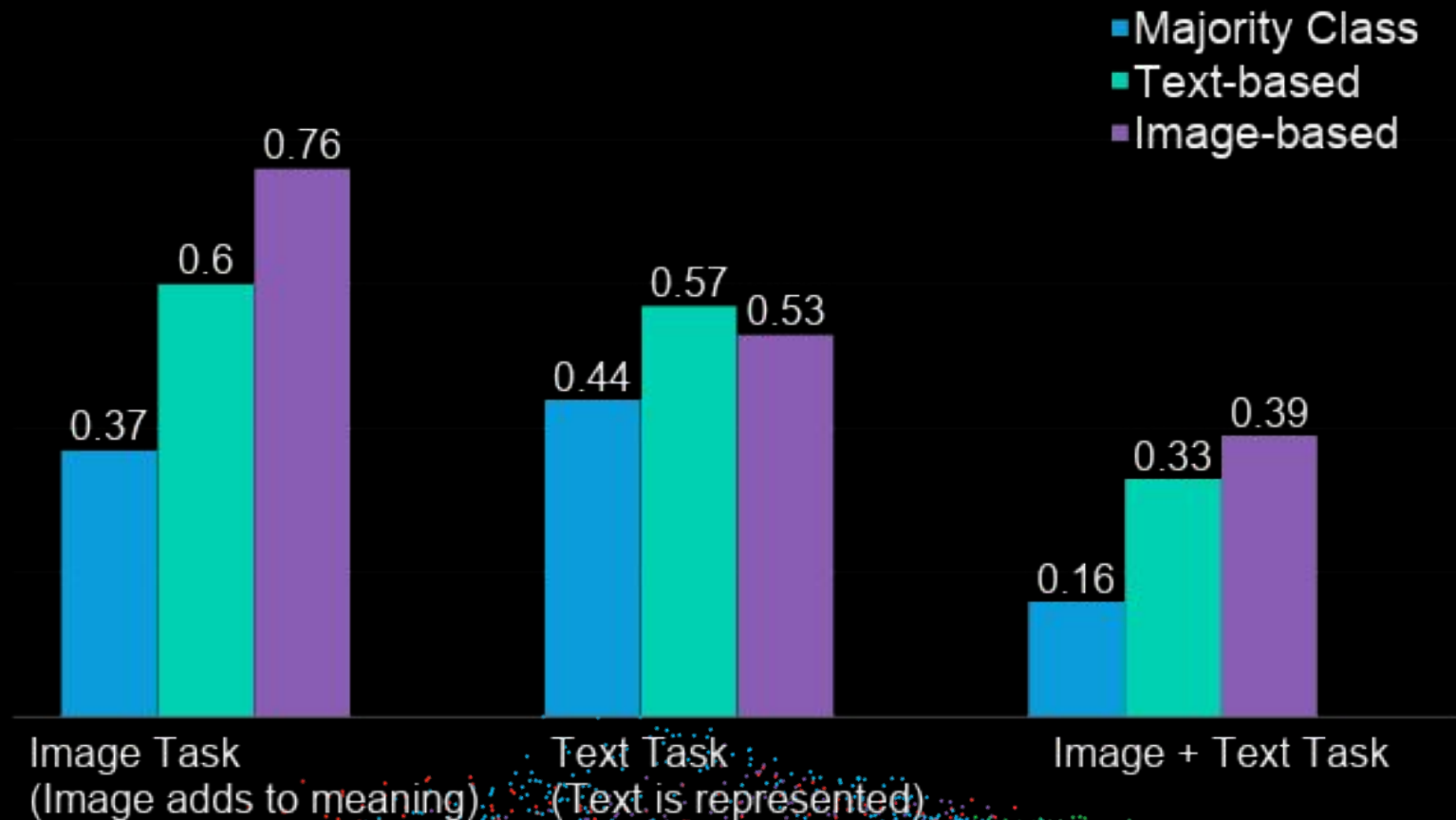


Bloomberg

Engineering

Prediction – Image-based Methods

- All tasks show improvements on the baseline
 - Image-based methods > Text-based on the Image Task
 - Image-based methods < Text-based on the Text Task
- Tuned InceptionNet is the best image-based method

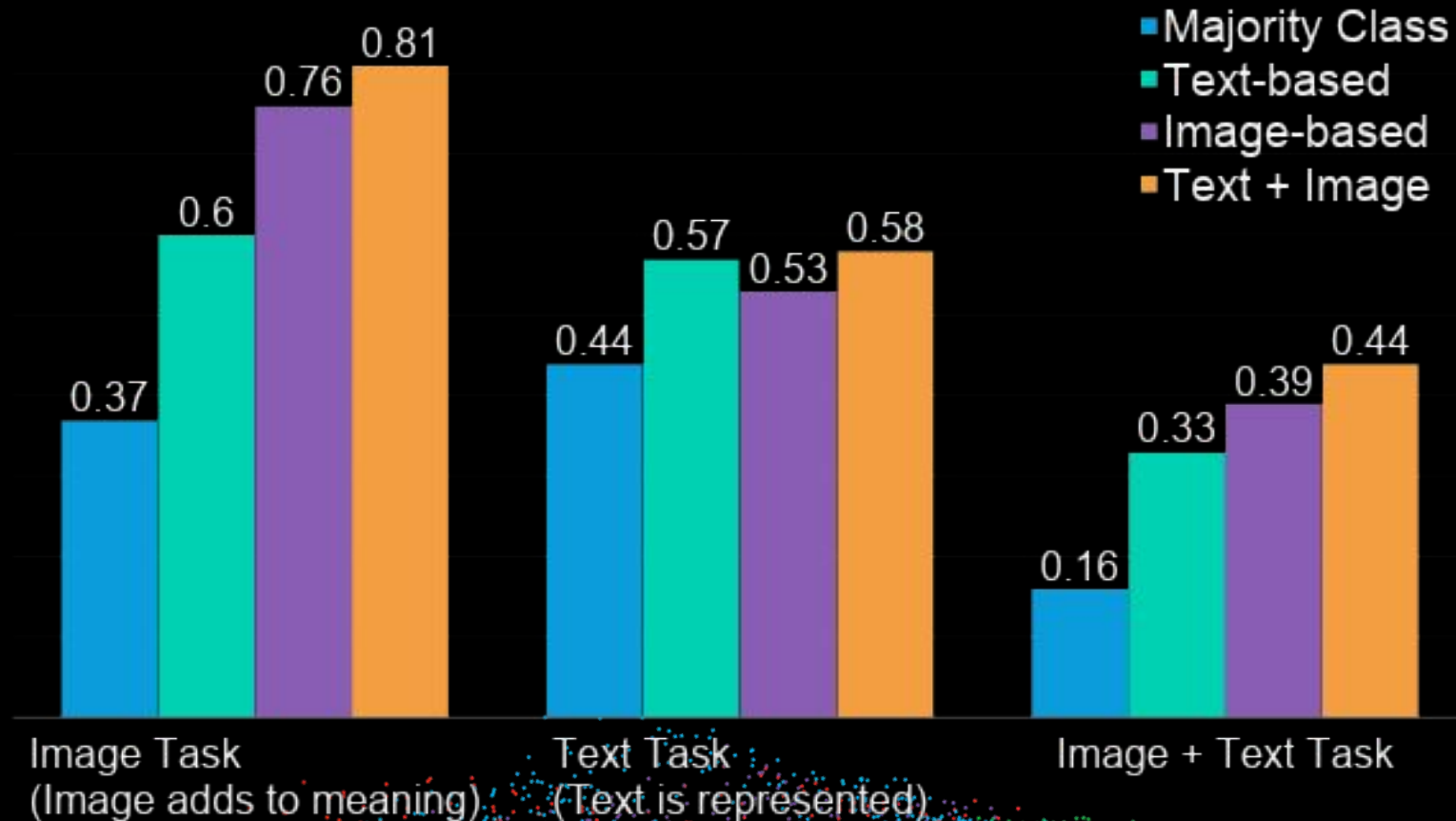


Bloomberg

Engineering

Prediction – Joint Text + Image Methods

- Improves over the best text-based or image-based methods
- LSTM + InceptionNet performs better
- Image task is much more predictable using models
- Text task is significantly harder (similar to humans)



Bloomberg

Engineering

Takeaways

Text-Image relationship in tweets is complex:

- Text does not always describe the image
- The image does not always illustrate text

Text-image relationship is likely useful for downstream applications

New classification schema and data set for text-image relationships on Twitter

- <https://github.com/danielpreotiuc/text-image-relationship>

Relationship type is predictable from both text and image

- Best results on each subtask are obtained by methods using different modalities (text or image)

Bloomberg

Engineering

Recent Related Work

- [Achieving Common Ground in Multi-modal Dialogue](#)
M Alikhani, M Stone - ACL 2020 Tutorial
- [Clue: Cross-modal Coherence Modeling for Caption Generation](#)
M Alikhani, P Sharma, S Li, R Soricut, M Stone - ACL 2020
- [Integrating Text and Image: Determining Multimodal Document Intent in Instagram Posts](#)
J Kruk, J Lubin, K Sikka, X Lin, D Jurafsky, A Divakaran - EMNLP 2019
- [CITE: A Corpus of Image-Text Discourse Relations](#)
M Alikhani, S Nag Chowdhury, G de Melo, M Stone - NAACL 2019

Analyzing Linguistic Differences Between Owner and Staff Attributed Tweets

Engineering

Bloomberg

Daniel Preoțiu-Pietro

Rita Devlin Marier

TechAtBloomberg.com

© 2020 Bloomberg Finance L.P. All rights reserved.

Motivation

User-level predictive tasks are very successful:

Age (Rao et al. 2010 ACL)

Gender (Burger et al. 2011 EMNLP)

Location (Eisenstein et al. 2010 EMNLP)

Personality (Schwartz et al. 2013 PLoS One)

Impact (Lampos et al. 2014 EACL)

Political Orientation (Volkova et al. 2014 ACL)

Mental Illness (Coppersmith et al. 2014 ACL)

Occupation (Preoțiuc-Pietro et al. 2015 ACL)

Income (Preoțiuc-Pietro et al. 2015 PLoS One)

Account Type (McCorriston et al. 2015 ICWSM)

User-level representations are used to improve results on a variety of tasks:

Sarcasm (Amir et al. 2016 CoNLL, Oprra & Magdy, 2019 ACL)

Comment moderation (Pavlopoulos et al. 2017 EMNLP)

Stance detection (Benton et al. 2018 EMNLP)

Motivation

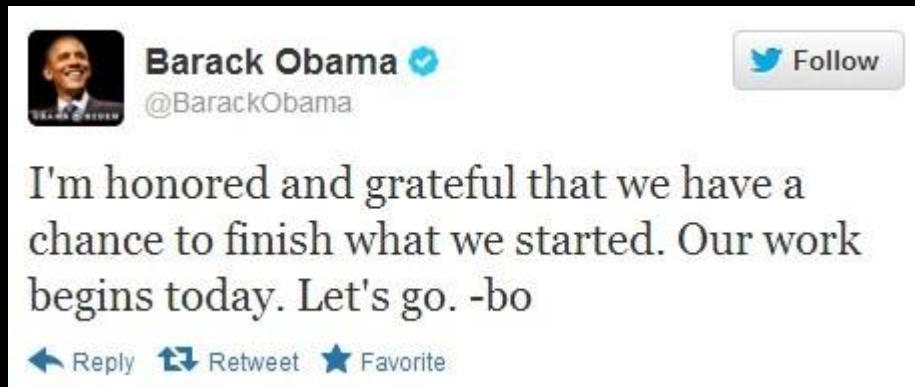
However, all these methods make a tacit assumption:

Tweets posted from an account are from a single person

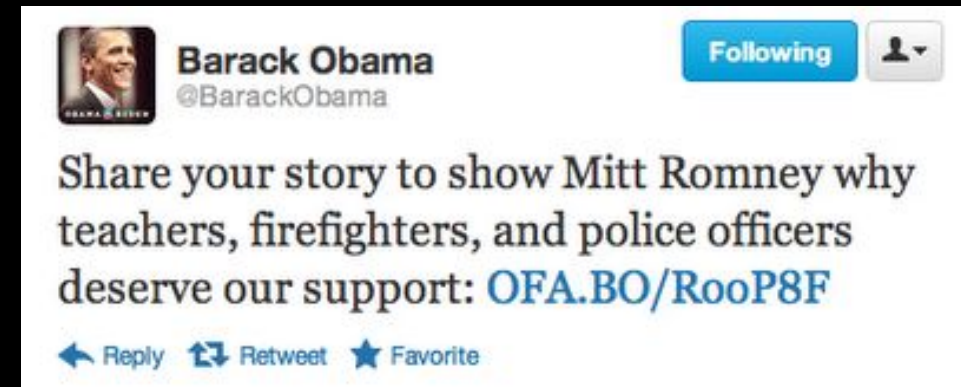
Motivation

One account = one person does not always hold

1. Politicians have staffers post their content



Signed by Barack Obama
Likely posted by Barack Obama



Not Signed by Barack Obama
Likely posted by staff

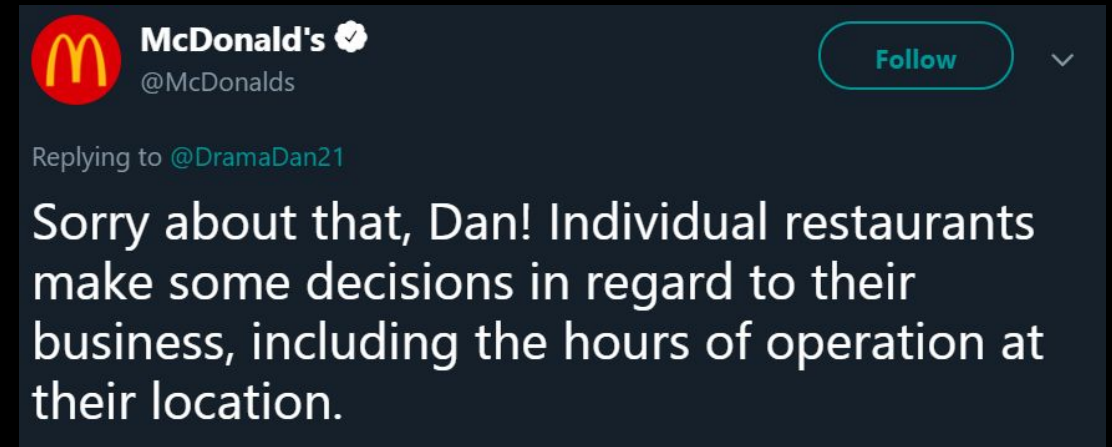
Motivation

One account = one person does not always hold

2. Companies use their account handles for marketing and customer support



Marketing tweet



Customer support tweet

Aim

The aim of our paper is to study differences between **types** of people posting from the same handle

Case study:

- Twitter
- U.S. politicians
- Owner vs. staff attributed

Data – Acquisition



Barack Obama ✓
@BarackObama

This account is run by Organizing for Action staff. Tweets from the President are signed -bo.

📍 Washington, DC
barackobama.com
🕒 Joined March 2007
📅 Born on August 4, 1961

 [Tweet to Barack Obama](#)

Data – Acquisition

Identify accounts which use a signature using a regex

- e.g., “tweets by me are signed ...”

Manually annotated

- If the account uses the convention
- The signature of the account (e.g., -bo)

Largest subgroup are U.S. politicians – 147 accounts

- We only use these accounts to control for the topic

Disclaimer:

- People may use the signature deceitfully
 - Albeit, little to gain and a lot to lose
- We refer to the task as author vs. staff **signed**



Data – Processing

Downloaded most recent 3,200 tweets from each account

- Retweets are removed

Search for signature in each tweet using a regex

- This is the label to predict
- If found, remove signature

Data set

- Size - 202,024 tweets in English
- Signed tweets - 4.8%
- Publicly available: <https://github.com/danielpreotiuc/signed-tweets>

Features

We experiment with traditional features to aid with our analysis

Tweet features

- **Length:** char, tokens
- **Type:** @-reply, URL
- **Time:** hour of day, day of week
- **Impact:** no. of retweets, no. of likes

Topics

- LIWC topics (Pennebaker et al. 1995)
- Word2Vec Clusters (Preoțiu-Pietro et al. 2015 ACL)

Sentiment & emotion predictions

- Positive or negative sentiment (Mohammad and Turney, 2013)
- Six Eckman emotions

Unigrams

Bloomberg

Engineering

Prediction

Binary classification task

Logistic Regression with Elastic Net regularization

Evaluated using ROC AUC (Area under the Curve)

- Data is class imbalanced (95 – 5)
- Random performance is 0.50

Experimental setups

Tweet-split Train



Test



User-split Train



Test

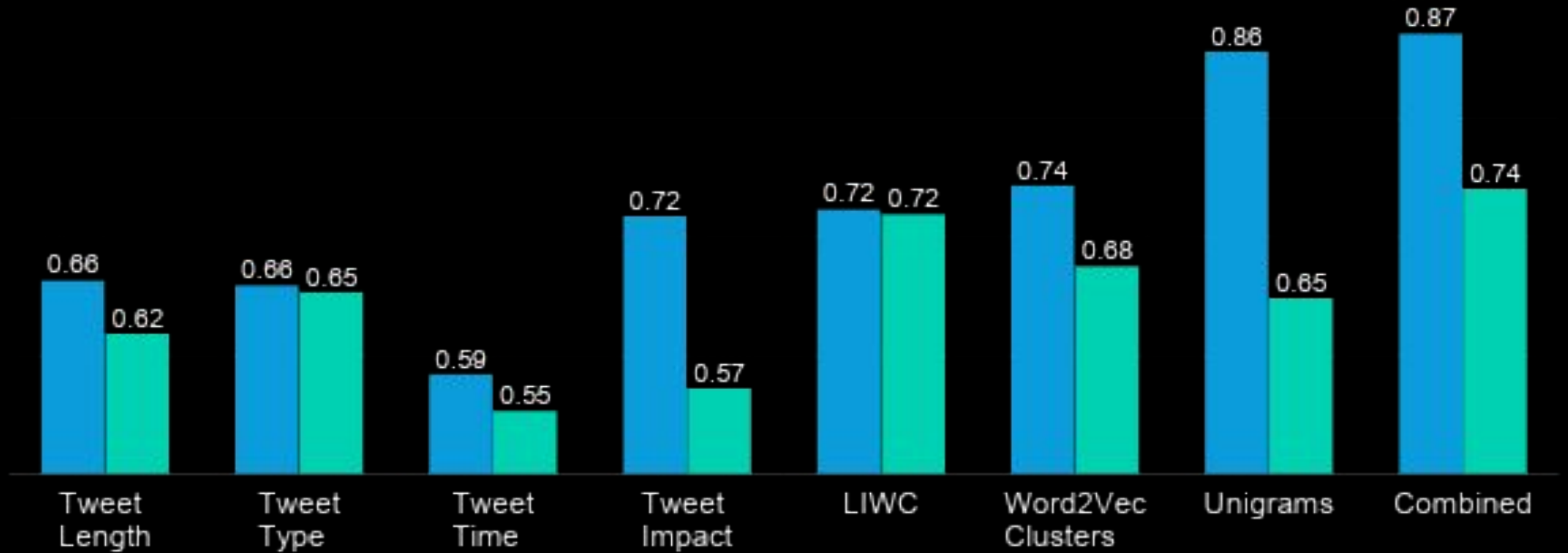


Bloomberg

Engineering

Prediction

■ Tweet-split ■ User-split



Analysis

Subsampled data from each account

- 1 signed – 9 unsigned
- Each account contributes at most 100 tweets

Aim is that no single user dominates

- the data set
- any label

Tweet Features – Mean Values

Feature	Owner	Staff
# Chars	105.4	102.4
# Tokens	23.2	21.4
Contains URL	45.7%	73.9%
@-Reply	4.2%	9.5%
Sent on Weekends	23.5%	20.7%
# Retweets	29.4	38.0
# Likes	82.3	79.1

* All differences between means shown in this table are significant at p .001, Mann-Whitney U test, Simes corrected

Analysis

- Signed tweets are more likely to be:
 - Longer
 - Sent on weekends
 - More liked, but less retweeted
 - Congratulations, condolences and support
 - More personal pronouns
 - More function words
 - More positive and negative sentiment
- No features are correlated with unsigned tweets:
 - More generic usage
- Other feature analysis in paper

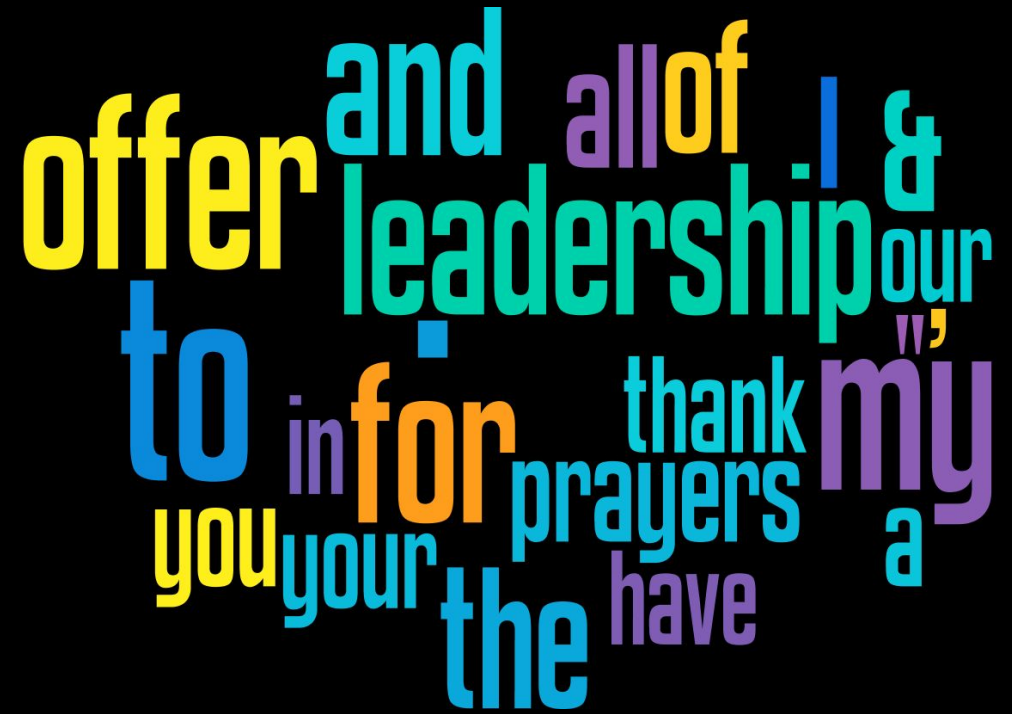
Tweet Features – Mean Values

Feature	Owner	Staff
# Chars	105.4	102.4
# Tokens	23.2	21.4
Contains URL	45.7%	73.9%
@-Reply	4.2%	9.5%
Sent on Weekends	23.5%	20.7%
# Retweets	29.4	38.0
# Likes	82.3	79.1

* All differences between means shown in this table are significant at p .001, Mann-Whitney U test, Simes corrected

Analysis

- Signed tweets are more likely to be:
 - Longer
 - Sent on weekends
 - More liked, but less retweeted
 - Congratulations, condolences and support
 - More personal pronouns
 - More function words
 - More positive and negative sentiment
- No features are correlated with unsigned tweets:
 - More generic usage
- Other feature analysis in paper



Unigrams

Analysis

- Signed tweets are more likely to be:
 - Longer
 - Sent on weekends
 - More liked, but less retweeted
 - Congratulations, condolences and support
 - More personal pronouns
 - More function words
 - More positive and negative sentiment
- No features are correlated with unsigned tweets:
 - More generic usage
- Other feature analysis in paper

PRONOUN
FUNCTION
SOCIAL
VERB
PREP
AFFECT

LIWC Topics

Takeaways

Not all tweets from a single account are posted by a single person

New data set released for research

- <https://github.com/danielpreotiuc/signed-tweets>

We are able to predict type of author with good precision

- Even for accounts unseen in training
- Different features transfer better to unseen handles

Use case results provide insight into the behavior of politicians

Thank You!

We are hiring for full-time positions:

- NYC – <https://careers.bloomberg.com/job/detail/86598>
- London – <https://careers.bloomberg.com/job/detail/85011>

Please reach out regarding summer 2021 research internships (NYC or London):

- <https://careers.bloomberg.com/job/detail/84208>

For more information about our work, research and programs:

<https://www.TechAtBloomberg.com/ai/>

Bloomberg

Engineering