GET BENT!

An Ideological Classifier

Steven Penn Markoe Data Scientist October 2020

VIEWER DISCRETION ADVISED

Twitter is a weird place.



President Trump's brother, Robert, has died. "It is with heavy heart I share that my wonderful brother, Robert, peacefully passed away tonight," he says in a statement. "He was not just my brother, he was my best friend. He will be greatly missed, but we will meet again."

10:29 PM · Aug 15, 2020 · Twitter for iPhone

426 Retweets 243 Quote Tweets 2.7K Likes







2020-08-16 02:55:33+00:00	Lmao at wrong trump trending
2020-08-16 02:55:30+00:00	The wrong trump died.
2020-08-16 02:55:17+00:00	The wrong trump died today.
2020-08-16 02:55:04+00:00	Y'all got wrong trump trending ðŸ~-ðŸ~-ðŸ~-ðŸ~-
2020-08-16 02:54:49+00:00	God took the wrong trump
2020-08-16 02:54:34+00:00	god that's twice now you took the wrong trump brother. couldn't you have just sent another flood?
2020-08-16 02:53:13+00:00	The wrong trump kicked the bucket ðŸ~″ðŸ¤-â€æ™,ï¸₪
2020-08-16 02:52:52+00:00	Dammit wrong trump
2020-08-16 02:52:37+00:00	@ god, wrong trump dude
2020-08-16 02:50:58+00:00	the wrong trump died but I'm sure Robert was an asshole too
2020-08-16 02:50:38+00:00	wrong trump died
2020-08-16 02:50:33+00:00	Noo Annabelle got the wrong trump đỹ~-
2020-08-16 02:49:55+00:00	I am not letting up on telling trump what I think of him and his immoral, corrupt, criminal regime. He has earned t
2020-08-16 02:49:44+00:00	To bad thr wrong trump was ill
2020-08-16 02:49:39+00:00	God killed the wrong trump #TrumpFailedAmerica
2020-08-16 02:49:37+00:00	wrong trump died
2020-08-16 02:49:10+00:00	you took the wrong trump đŸ¥′
2020-08-16 02:48:34+00:00	wait wrong trump
2020-08-16 02:48:23+00:00	Simpson's predicted the wrong trump smh. Jk I wish death upon nobody not even my worst enemy
2020-08-16 02:48:12+00:00	the wrong trump diedmy wife just said you could eliminate the National Debt by selling tickets to dit's funei
2020-08-16 02:47:45+00:00	The wrong trump died. smdh
2020-08-16 02:47:20+00:00	The wrong trump died ô'y"'
2020-08-16 02:47:02+00:00	Ffs wrong trump đŸ¤¦đŸBâ€ā™,ï ₫Ÿ~,
2020-08-16 02:46:44+00:00	They really took the wrong trump $\delta \ddot{V}$
2020-08-16 02:46:31+00:00	I fcking don't care wrong trump died as far as I'm concerned
2020-08-16 02:46:21+00:00	damnit. the wrong trump diedôŸ~'
2020-08-16 02:46:06+00:00	I think the wrong trump died.
2020-08-16 02:46:00+00:00	Got the wrong trump
2020-08-16 02:45:49+00:00	
	that was the wrong trump @ karma
2020-08-16 02:45:21+00:00	I don't know much about Robert Trump but, to be honest the wrong trump died. Sorry not sorry.
2020-08-16 02:44:31+00:00	dear god, u took the wrong trump family member
2020-08-16 02:43:15+00:00	the wrong trump died smh
2020-08-16 02:42:39+00:00	damn cuh y'all took the wrong trump ðÝ™
2020-08-16 02:40:17+00:00	yo god wrong trump bro
2020-08-16 02:39:36+00:00	they got the wrong trump
2020-08-16 02:39:19+00:00	Not the wrong trump dying 🙄
2020-08-16 02:38:21+00:00	Dammit! The wrong trump died! ỡŸ¤¦â€â™€ï,⊠
2020-08-16 02:38:05+00:00	The wrong trump died
2020-08-16 02:37:39+00:00	anabelle baby wrong trump ðŸ~"ðŸ~"
2020-08-16 02:37:34+00:00	I might get suspended for this, but The wrong trump died.
2020-08-16 02:36:32+00:00	Damn the wrong trump died ðŸ~"
2020-08-16 02:35:36+00:00	Damn wrong trump :/
2020-08-16 02:34:31+00:00	@ god you took the wrong trump this time δΫ'€
2020-08-16 02:34:18+00:00	The wrong trump died
2020-08-16 02:34:11+00:00	The wrong trump died

Problem Statment

 Looking at a politically charged Twitter hashtag, how do I know which users are conservatives and which are liberals? Given any polarizing topic, can I classify which conflicting ideology the individuals represent?

Some Assumptions

1. Political twitter is dominated by political users.

*https://www.pewresearch.org/politics/2 019/10/23/national-politics-ontwitter-small-share-of-u-s-adultsproduce-majority-of-tweets/ 2. Political Twitter is just two competing ecosystems.

^American politics

Scraping

- GetOldTweets3 and snscrape
- All tweets, hashtags
- No retweet content

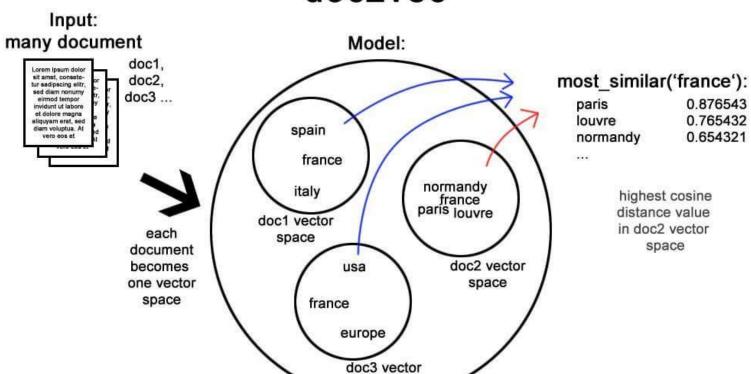
- Suspended users
- Name changes
- Bots
- Hundreds of thousands of individual tweets

Data Overload!

Count Vectorizer? or Tf-Idf?

"You shall know a word by the company it keeps." -J.R. Firth

doc2vec



space

RedTeam!









BlueTeam!

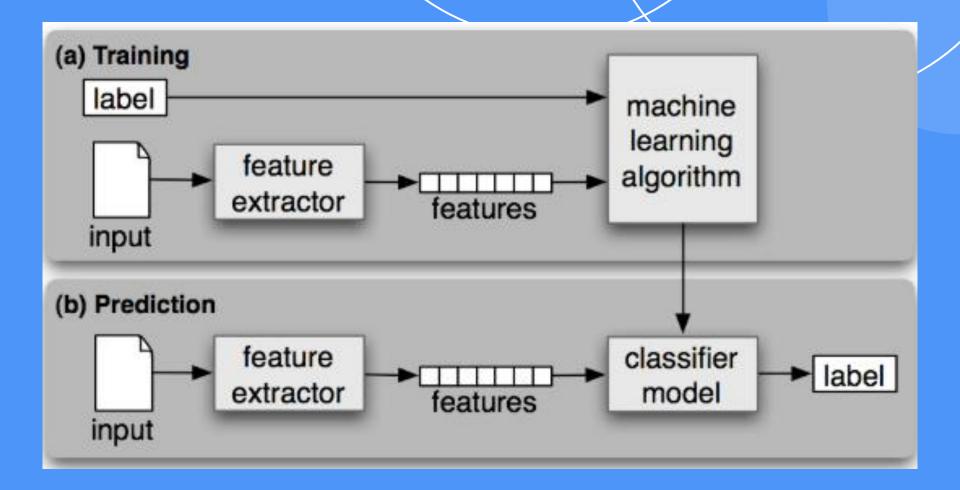












Modeling

Doc2Vec

- 300Vectors
- 300 min
- 40 epochs

SVC

- Nothing
 - Nearly perfect accuracy

CONCLUSIONS!

CONCLUSION 1

With the appropriate training set and Doc2Vec, users participating in a 'barbershop dispute' can be classified with relative ease.

CONCLUSION 2

Twitter is an enigma wrapped in two more enigmas.

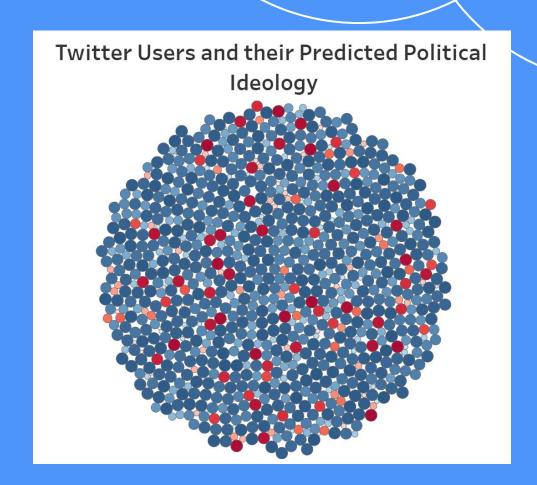
Conclusion 3

Training Data absolutely includes inherent bias.

Next Steps!

getbent.com: Discover the political ideologies of your brand's followers!

Tableau!



Questions?

Thanks so much!

Steven Markoe steven.markoe@gmail.com