Classifying Tweets from Russian Trolls

Joey Goodman

September 12, 2019

Russia's disinformation campaigns, then and now

The New York Times

Russian 2016 Influence Operation Targeted African-Americans on Social Media

Yevgeny Prigozhin, left, and the Russian leader Vladimir V. Putin, center, at a dinner in 2011. Mr. Prigozhin was indicted by American prosecutors for his involvement in interfering in the 2016 presidential election. Pool photo by Misha Japaridze

"I hope this is not the new normal, but I fear it is."

Robert Mueller, on Russian election meddling (July 24, 2019)

Examples of Russian troll content on Twitter

Research to date on Russian twitter accounts

Linvill and Warren (2019)

- Collected tweets from 2,800 Russian troll accounts after they were banned
- Introduced a taxonomy for different types of Russian accounts

Im et al. (2019)

- Built a model to make predictions at the *account-level*
- Collected control accounts from a historical sample of Twitter data

My contribution:

- Make predictions at the *tweet-level*
- Construct a more representative control group

Data overview

Right Troll tweets

3 million tweets

- Tweeting between ‘14 and ’18
- Excluding retweets
- Collected by Linvill & Warren (Clemson University)

Data overview

Right Troll tweets

3 million tweets

- Tweeting between ‘14 and ’18
- Excluding retweets
- Collected by Linvill & Warren (Clemson University)

Verified tweets

170,000 tweets

- Queried using the 80 most common hashtags appearing in the Right Troll dataset

Data overview

Right Troll tweets	Verified tweets
170,000 tweets	**170,000 tweets**
• Down-sampled to match shape	• Queried using the 80 most common hashtags appearing in the Right Troll dataset

Data overview

Right Troll tweets

170,000 tweets

+

Verified tweets

170,000 tweets

→

Final dataset

340,000 rows

4 feature matrices

Words

america	country
#	#

Emojis

😄	👍
#	#

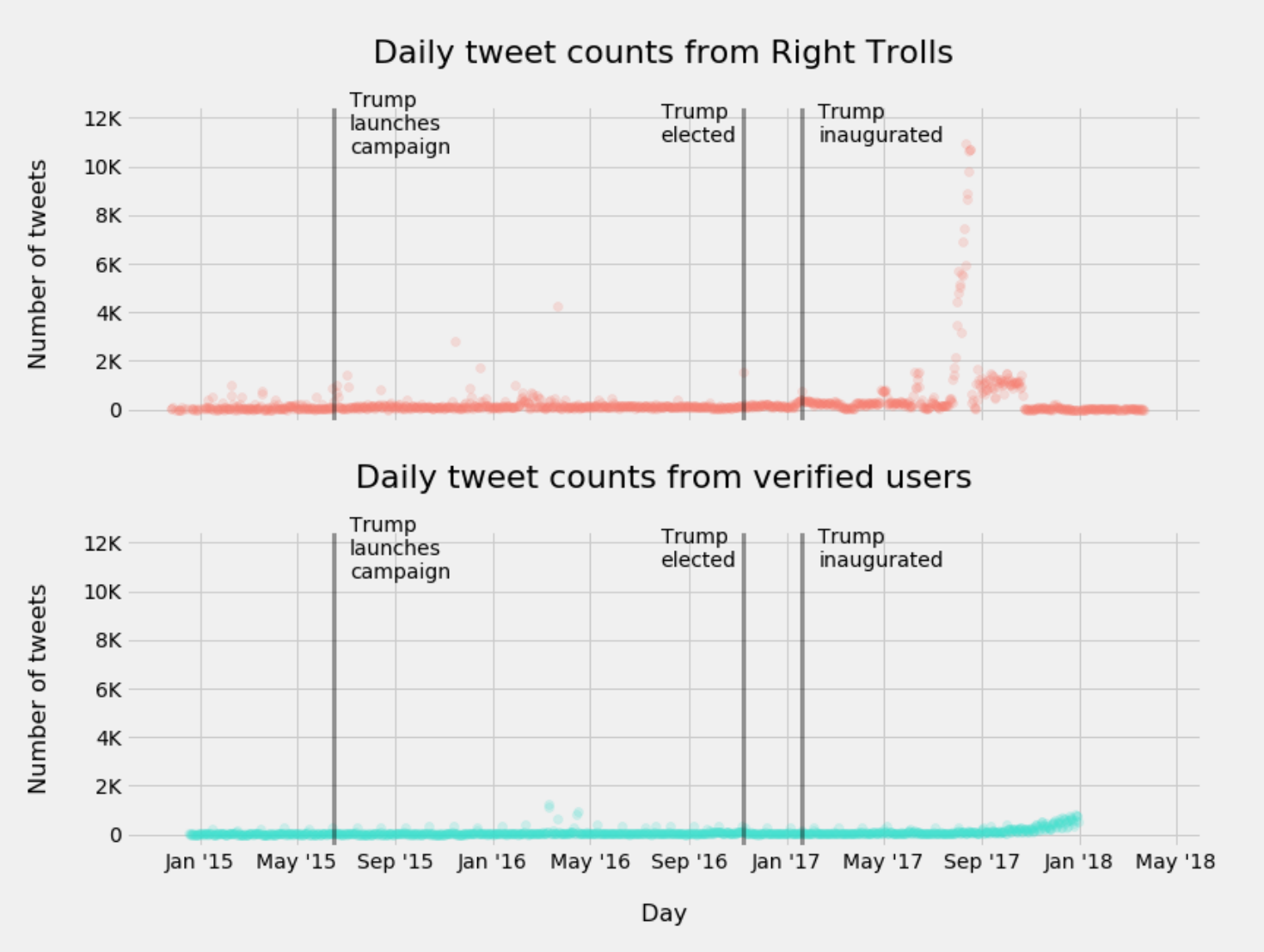
Hashtags

#Trump	#MAGA
#	#

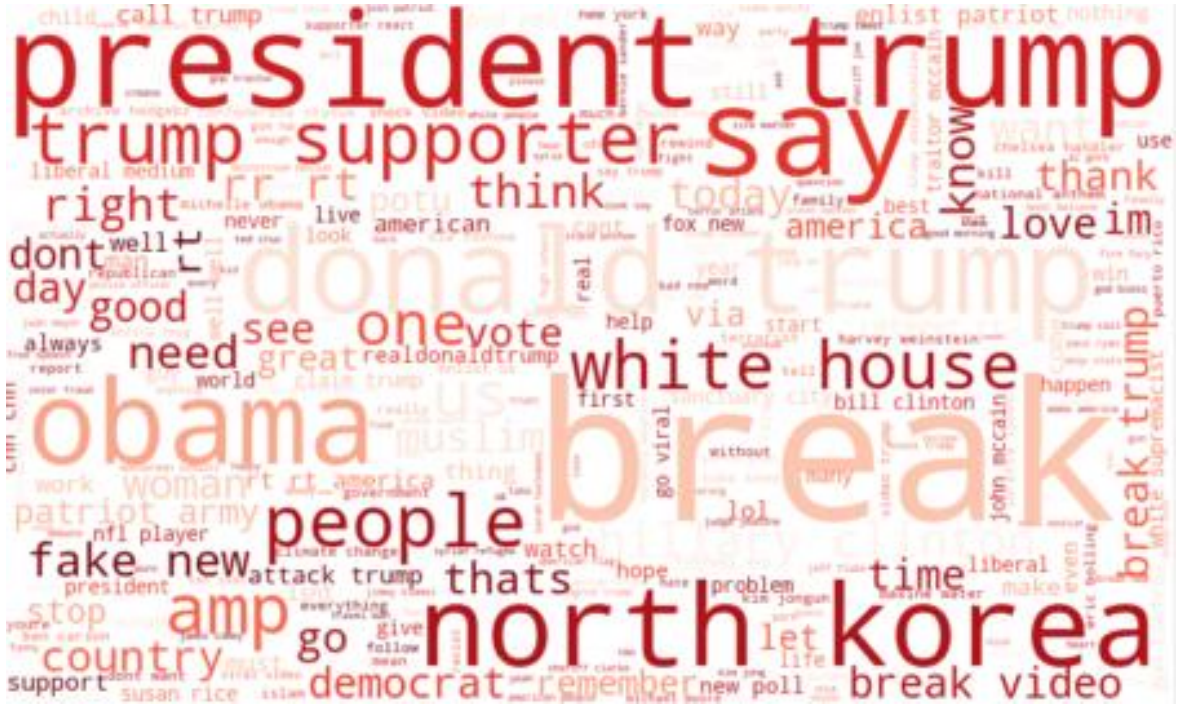
Numeric features

links	pics
#	#

Visualizing tweeting patterns

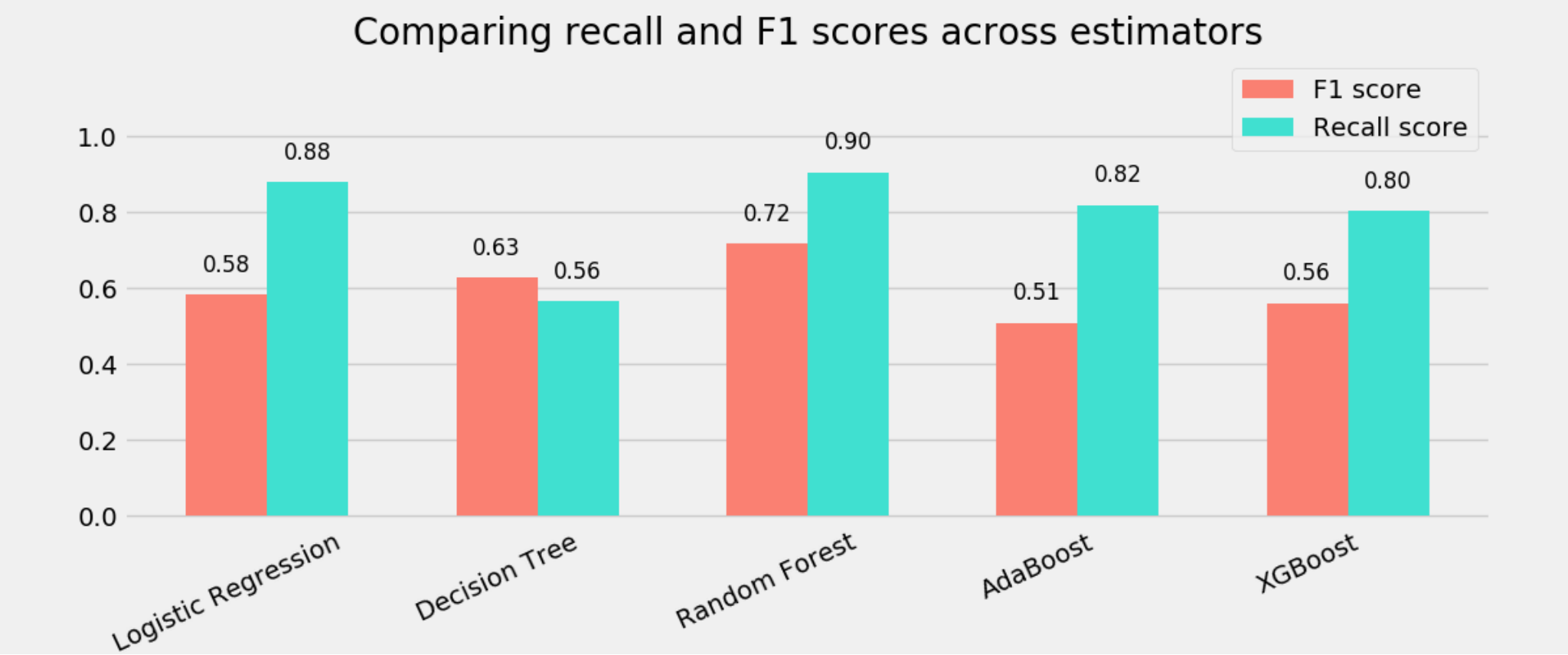


Troll vocabulary vs. verified user vocabulary



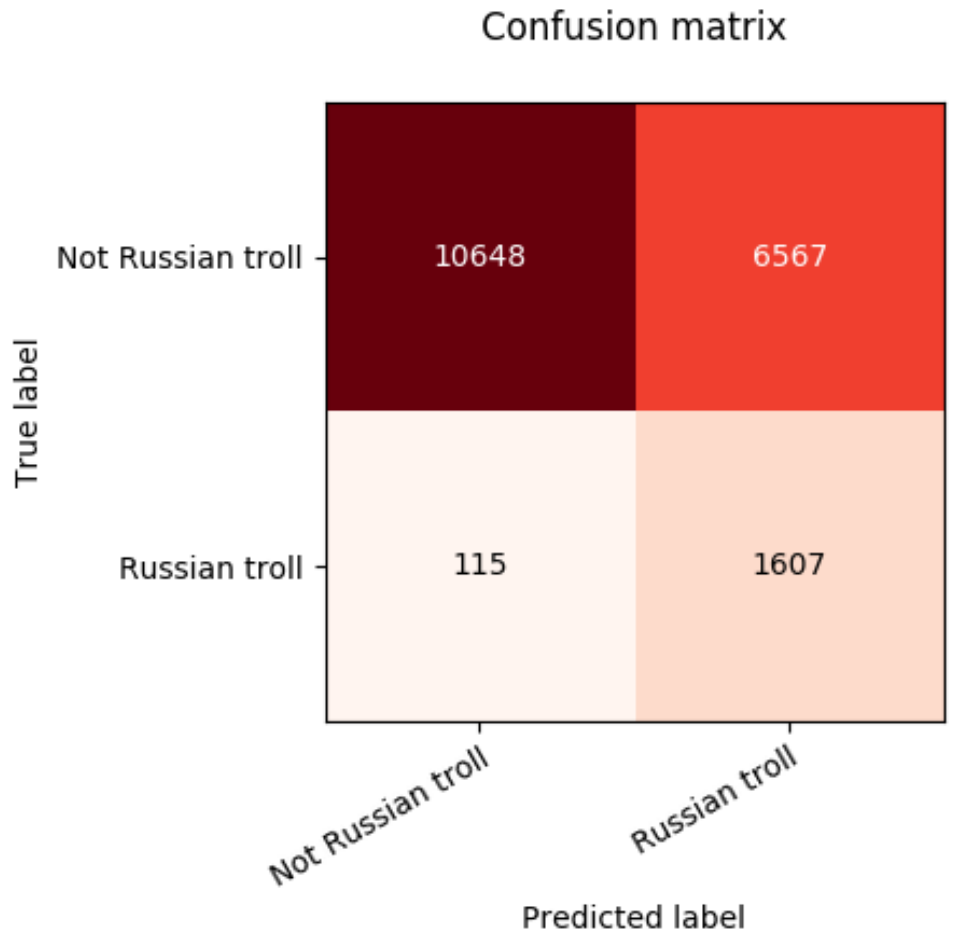
Comparing results from various classifiers

Scoring metric: **Recall** (trying to minimize false negatives)



Best estimator: Random Forest

Best recall with grid search: **93.3%**



Top 10 most important features

- pics_count
- trump
- pct_upper
- hashtags_count
- rt
- break
- #trump
- happy new
- mentions_count
- video

Examples of misclassified tweets

@TEN_GOP

Muslims protesting today in NYC. When will we see them taking over the streets in a movement against terrorism. #BodegaStrike https://t.co/gOU2obcHXV

2017-02-02 21:41:00

@RYANAMBISHOP

#StopIslam #IslamKills #PrayForBrussels please stay safe

2016-03-22 19:02:00

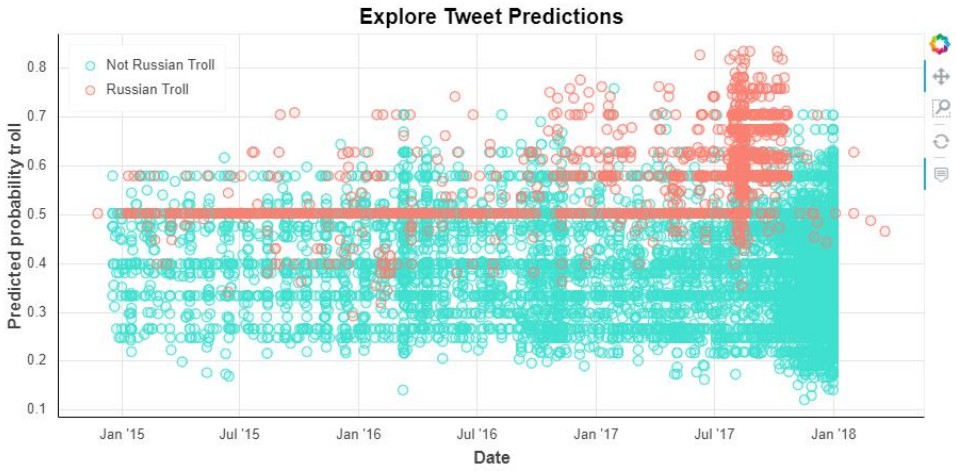
@COVFEFENATIONUS

@ChuckyStorms @john_mackay13 @JasperAvi @jaketapper @IronStache @TheDemocrats @SenateDems @HouseDemocrats @CNN @CNNI @CNNPolitics @CNNSitRoom @WolfBlitzer @JakeTapper @TheLeadCNN @BrianStelter @AnaNavarro @DonLemon @VanJones68 @AndersonCooper @AC360 @JimAcosta CNN IS #FAKENEWS! CNN IS #FAKENEWS! CNN IS #FAKENEWS! CNN IS #FAKENEWS! CNN IS #FAKENEWS! CNN IS #FAKENEWS! #FAKENEWS #MAGA

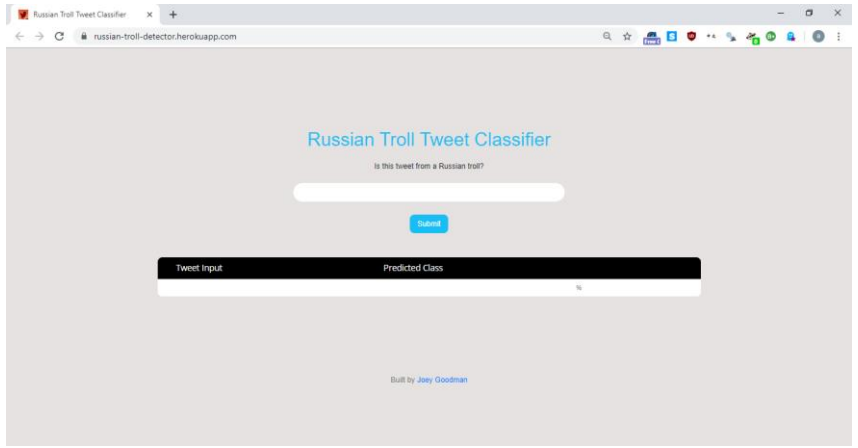
2017-12-02 02:13:00

Explore the data

Visualizing tweet predictions



Web app to make new predictions



Future work

- Test other word vectorizing approaches
- Further tune models
- Explore word embeddings, to help with reducing dimensionality
- Collect more verified tweets

Thank you!

Find me @ joeygoodman.us

Code + visuals: github.com/yontartu/bot-vs-human

Sources

Russian Troll tweets dataset: FiveThirtyEight

Verified tweets dataset: collected using twint

New York Times, "Russian 2016 Influence Operation Targeted African-Americans" (December 17, 2018)

New York Times, "Highlights of Robert Mueller's Testimony to Congress" (July 24, 2019)

Twitter, "Update on Twitter's review of the 2016 US election" (January 31, 2018)