AUTOMATIC DETECTION OF EXTREME POLITICAL GROUPS

Natural Language Classification of Reddit Subforums

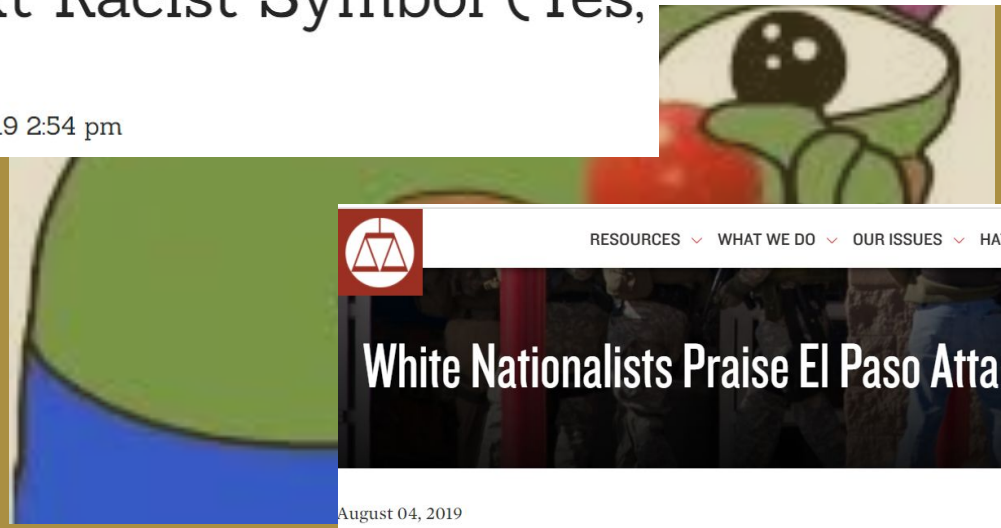
After 9 months, Reddit finally bans group spreading thinly veiled anti-Semitism

Community had grown to 60,000-strong, and was prompting concerns elsewhere on social media platform; similar,

No Joke: White Nationalists Are Now Using Clowns To Spread

White Nationalists Adopt Clowns as Their Next Racist Symbol (Yes, Seriously)

By Jared Holt | April 4, 2019 2:54 pm



RESOURCES WHAT WE DO OUR ISSUES HATEWATCH HATE MAP PODCAST DONATE

White Nationalists Praise El Paso Attack and Mock the Dead

August 04, 2019

by Michael Edison Hayden

The Guardian

one, funded by readers

Search jobs

Reddit bans Qanon subreddits after months of violent threats

Reddit was key to Qanon's emergence from the fringe corners of the internet and into more mainstream conspiracy and far-right media circles.

2018

Trump forum

How Charlottesville forced Reddit to clean up its act

Reddit CEO Steve Huffman clarifies that racism is not welcome on the platform

'I try to stay neutral on most political topics, but this isn't one of them.'

26 October 2017

Racism and anti-Semitism surged in corners of the Web after Trump's election, analysis shows

shooting fiasco

By Hayley Tsuka

Reporter

June 16, 2016 at 12:30 p.m.

white supremacist alt-right forum

Users weren't happy.

By Aja Romano | @ajaromano | Nov 16, 2016, 1:20pm EST

Frens I just got kicked out of my 109th subreddit... may I join yours? (i.redd.it)
submitted 14 days ago by Furscist SchutzFren
198 comments share save hide give award report crosspost [l+c]

A Broad View

Goal:

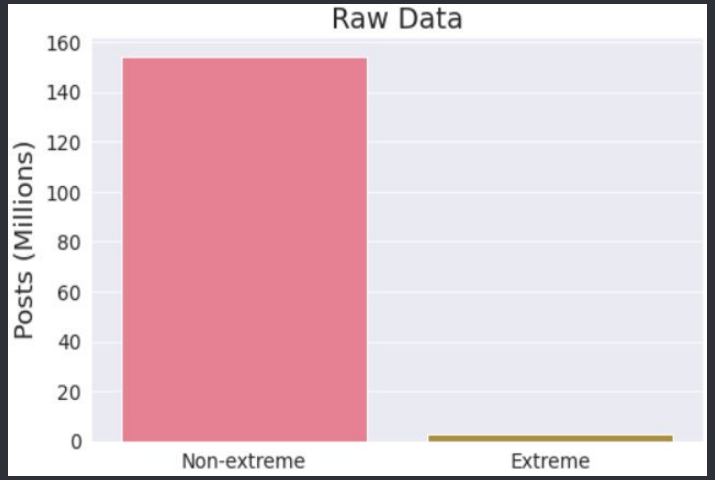
- Auto-flag likely groups
- Notice groups while still small
- Get in front of public criticism

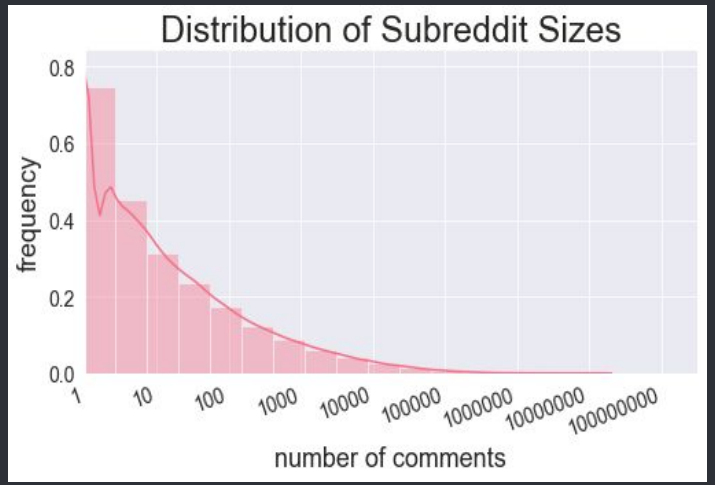
Who To Look At:

- Openly Political
- Media Scrutiny
- Banned Subreddits

Data

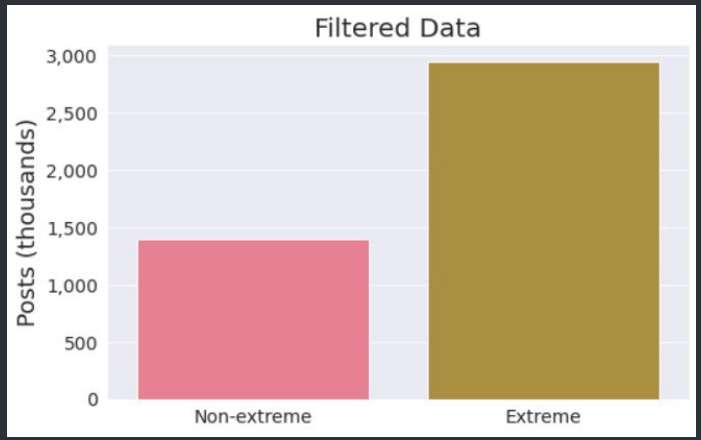
- August & September, 2017
- 85 million posts
- 98 thousand boards





Data

- 3156 boards examined
- 20 classified “extreme”
- 44,000 posts
- Cleaned and grouped to 100s of posts

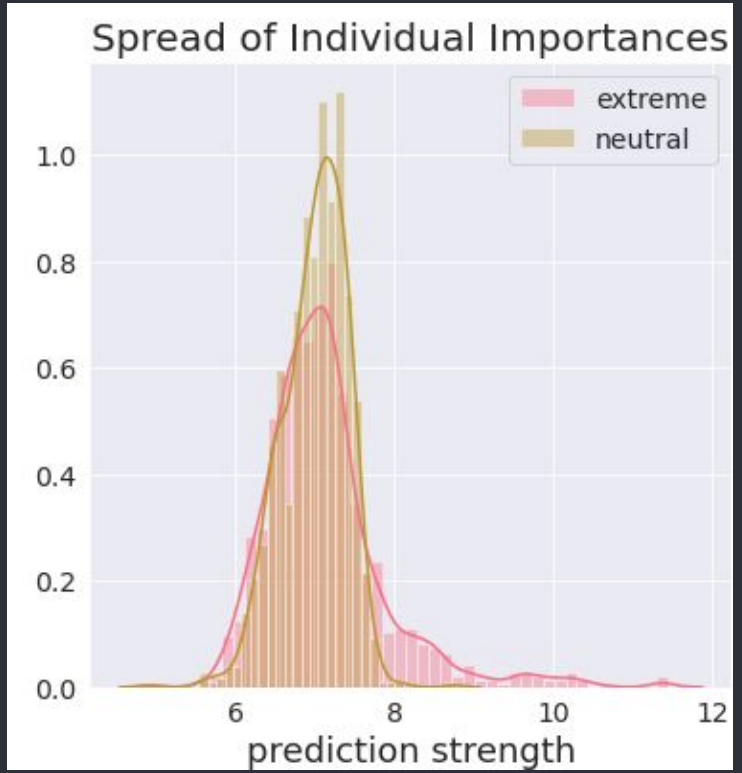


Making the Predictions

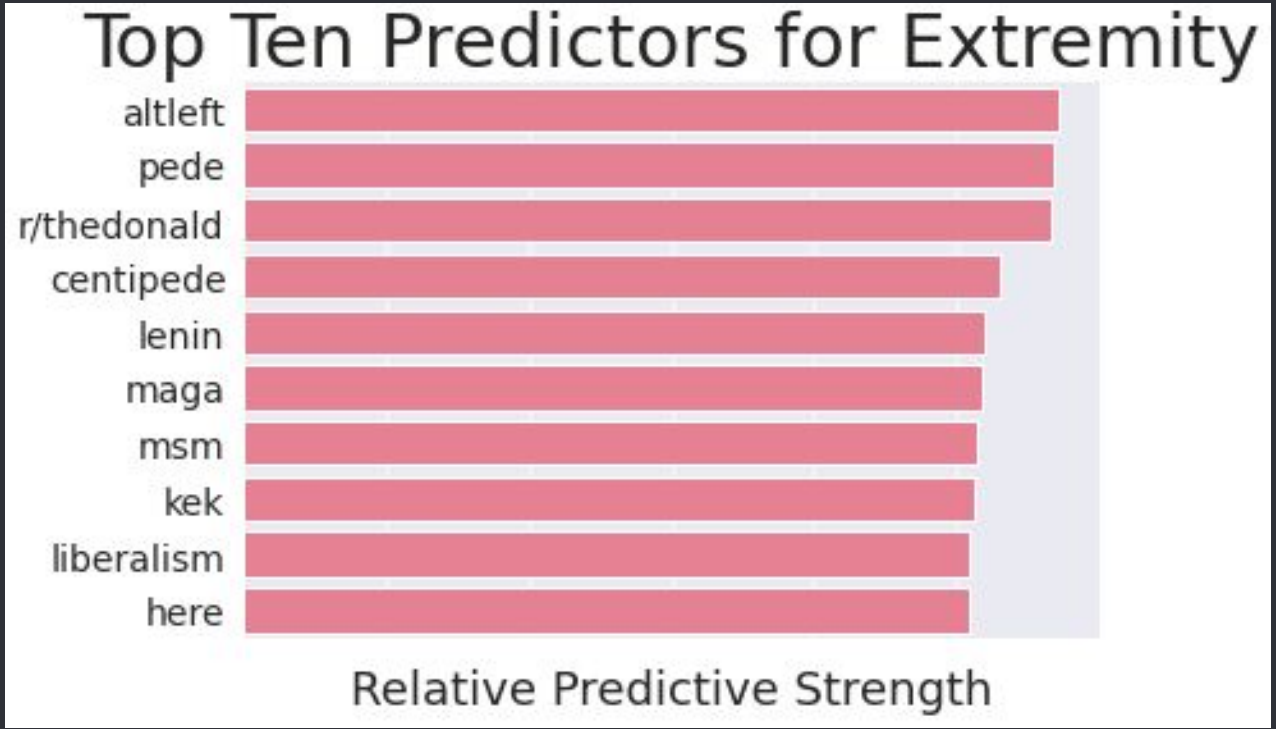
1. Group of 100 posts

2. One-or-Two Word Phrase Frequencies

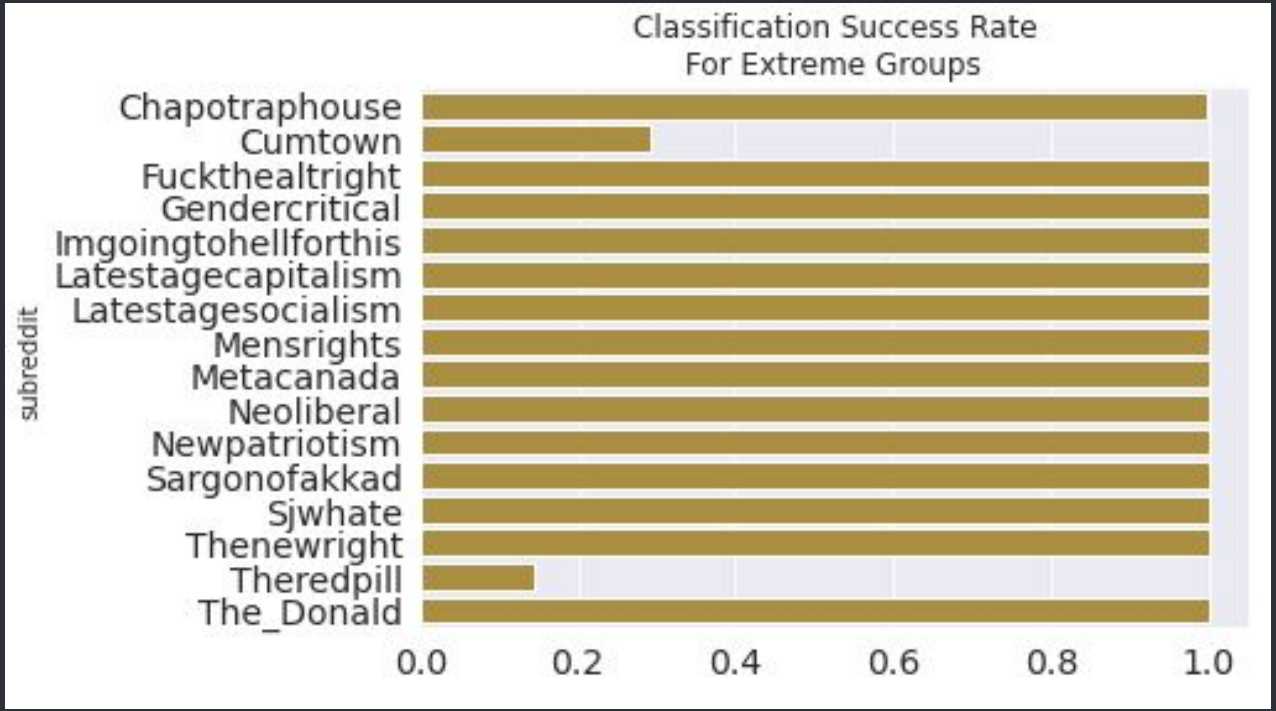
3. Predictions on ***not*** seeing phrase



- Making the Predictions



Making the Predictions



Limitations & Concerns

- “Sometimes the Cure Is Worse Than The Disease"
- Data quality
- Unclear generalizability
- Time-boundedness
- Main-use-case applicability (do our basic assumptions hold)?

Moving Forward

- More Sophisticated Modelling
- Real-World Testing
- Extend Data
- Limited Rollout

Thank you for listening!

Contact Info

James Shaw

Email: shawjr3@gmail.com

Github: godelayheehoo

Linkedin: James-shaw-848984104

Twitter: @Pearls_Before_Quine