Predicting Future Political Ideology from Present Self-Description:
A Study of US Twitter Users, 2015-2020

Rajrishi Das, Jason Jones PhD (advisor)

Introduction

- Every new Twitter user is prompted: "**Describe yourself in 140 characters or less.**" Their response (their "**bio**") is public-facing and freely editable; users typically update their bio roughly once per year. Studying such bios at scale yields important insight into trends in *personally expressed identity*.
- Recent work, for instance, shows that **since 2015, Americans have been adding political keywords (e.g. 'progressive', 'MAGA') to their bios more rapidly than keywords of any comparably major category (e.g. sports or religion).** [1]
- This study aims to further this research by **determining how accurately the political** <u>ideologies</u> **expressed by Americans** <u>in 2020</u> **can be predicted from the text of their** <u>previous</u> **self-descriptions –** **that is, from the text of their Twitter biographies from each year between 2015 and 2019.**
- Furthermore, it seeks **to identify notable year-over-year trends in token-level prediction** by identifying the words whose inclusion in a user's self-description in any given year strongly predicts their future expression of conservative or liberal partisanship.
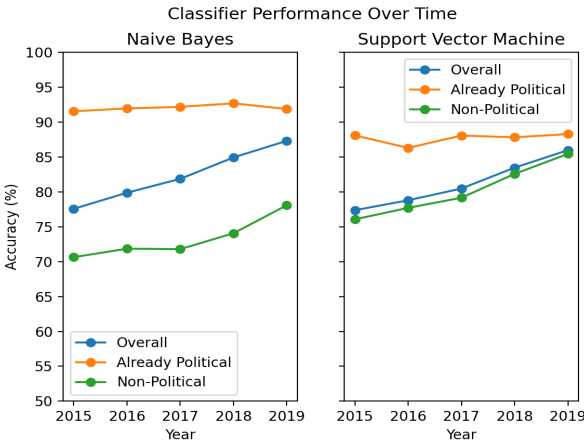
Data and Methods

Data: The bios of ~45k Twitter users as recorded once each year from 2015-2020. Each user is labelled as 'liberal' or 'conservative' based on their inclusion, in their 2020 bio, of any of the following keywords:
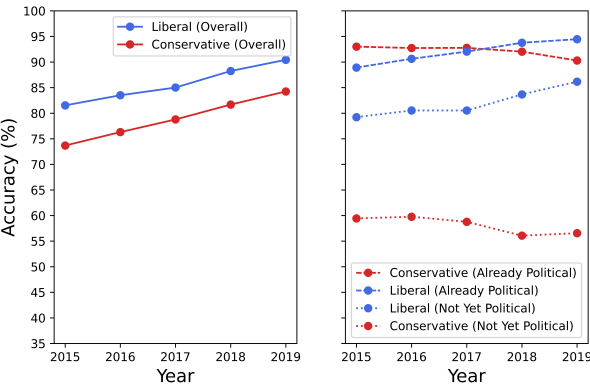
- **Liberal:** liberal, progressive, democrat, socialist, voteblue, blm, blacklivesmatter, feminist, freepalestine, feelthebern, bidenharris, biden
- **Conservative:** conservative, republican, libertarian, trump, trumppence, maga, kag, deplorable, backtheblue, bluelivesmatter, alllivesmatter, wwg1wga, qanon, nra, pro-life

Methods: **Naïve Bayes** and **Support Vector Machine** text classifiers were trained on a balanced (50% conservative, 50% liberal) random subset (n=~25K) of the above data (consisting only of bios from 2015-2019 and their labelled 'ideology'), and tested on a different random subset (n=~6K).

Results

Classifier Performance Over Time

Naive Bayes Classifier Performance on Liberal and Conservative Bios

New* Predictors by Year: ** For the year 2015, table lists the 10 tokens from 2015 bios with the highest predictive power (weighted by prevalence). For years 2016-2019, lists tokens from the 100 most predictive tokens that were not among any of the previous years' 100 most predictive tokens.*

$$Predictive\ Power\ (Liberal) = \frac{Pr(token \in bio\ that\ is\ Liberal\ by\ 2020)}{Pr(token \in bio\ that\ is\ Conservative\ by\ 2020)}$$

$$Predictive\ Power\ (Conservative) = \frac{Pr(token \in bio\ that\ is\ Conservative\ by\ 2020)}{Pr(token \in bio\ that\ is\ Liberal\ by\ 2020)}$$

References
1. Rogers, Nick and Jason J Jones. 2021. "Using Twitter Bios to Measure Changes in Self-Identity: Are Americans Defining Themselves More Politically Over Time?" *Journal of Social Computing* 2(1):1–13.

Year	Liberal	Conservative
2015	feminist, ig, writer, music, snapchat, student, university, lover, artist, #blacklivesmatter	god, christian, father, family, husband, country, fan, christ, sports, american
2016	intersectional, she, ya, opinions, ain't, write, black, rip, work, actor, inquires, queen	time, people, get, account, show, nascar, independent, vet, org, company, never, first
2017	#blm, #resist, alumna, her, 🏳️‍🌈, ally, views, teacher, advocate, reader	co, make, president, 🇺🇸, constitutionalist, #prolife, beautiful, help
2018	twitch, mom, he, icon, 🌹, him, educator, nonprofit, account, they, blm	#kag, twitter, friends, lord, #draintheswamp
2019	hers, advocacy, libra, #resistance, 📿, bi, socialist	#wwg1wga, #patriot, #buildthewall, deplorable, reagan, #qanon, day, supporter