

YICHEN HU

ANALYSIS OF ANDREW YANG SUPPORTERS ON REDDIT

PROJECT GOAL

- ▶ Andrew Yang is a 2020 democratic candidate for president who does not have the name recognition of more well known candidates
- ▶ Many of the ways of addressing this are either expensive (ex: buying ads) or have low success rates (ex: phone banking)
- ▶ An approach that could have a higher conversion rate would be to identify people who are more likely to be interested in Andrew Yang and then having conversations with them
- ▶ I hope to assist in that by identifying interests of current Yang supporters with the thought that others who share those interests but are not current Yang supporters would be more likely to become a Yang supporter than the rest of the general population

DATA GATHERING

- ▶ Reddit data was gathered using the PRAW library
- ▶ First, I identified users that had recently posted on the Andrew Yang subreddit, subreddits of other 2020 candidates and subreddits supporting the democratic party
- ▶ For each user, I then gathered their recent comments and what subreddits those were made on
- ▶ For the most common subreddits in my scrape, I exported those to a list, kept the ones that were relevant outside of Reddit, grouped what I could and turned those into features
- ▶ In total, 3.54 million individual comments were scraped from over 18,000 subreddits

MODEL RESULTS – SUBREDDIT BASED FEATURES ONLY

- ▶ Logistic regression classifier
- ▶ Baseline score: .7368 (majority class = not Yang supporters)
- ▶ Accuracy score: .7695
- ▶ This heavily predicted the majority class which isn't that useful for me
- ▶ Used Synthetic Minority Over-sampling Technique (SMOTE) to artificially create more minority class observations
- ▶ This decreases my accuracy score (to .6715) but increased my sensitivity from .2222 to .5778

TOPICS AND INTERESTS OF YANG SUPPORTERS

► Top 10 interests/topics most associated with Yang supporters

- | | |
|---------------------------|--------------------|
| 1. Universal Basic Income | 6. Futurology |
| 2. Asian culture | 7. Jordan Peterson |
| 3. Sam Harris | 8. Hip-hop |
| 4. Cryptocurrency | 9. Android |
| 5. Libertarians | 10. Teenagers |

► Top 5 interests/topics least associated with Yang supporters

1. Cars
2. Environment
3. Liberal Gun Owners
4. Ask Gay Bros
5. Documentaries

MODEL RESULTS – NLP FEATURES ONLY

- ▶ I wanted to avoid looking at all comments so I focused just on the r/politics subreddit since that was a frequent posting spot for both Yang and non-Yang supporters
- ▶ This model is not as applicable as my previous model since not everyone posted in r/politics
- ▶ Baseline: .7590
- ▶ Logistic Regression Classifier with TF-IDF vectorizer
 - accuracy: .8208
 - sensitivity: .3243
- ▶ Using SMOTE, sensitivity increased to .7162

WORDS / PHRASES OF YANG SUPPORTERS

- ▶ Top 10 words/phrases most associated with Yang supporters

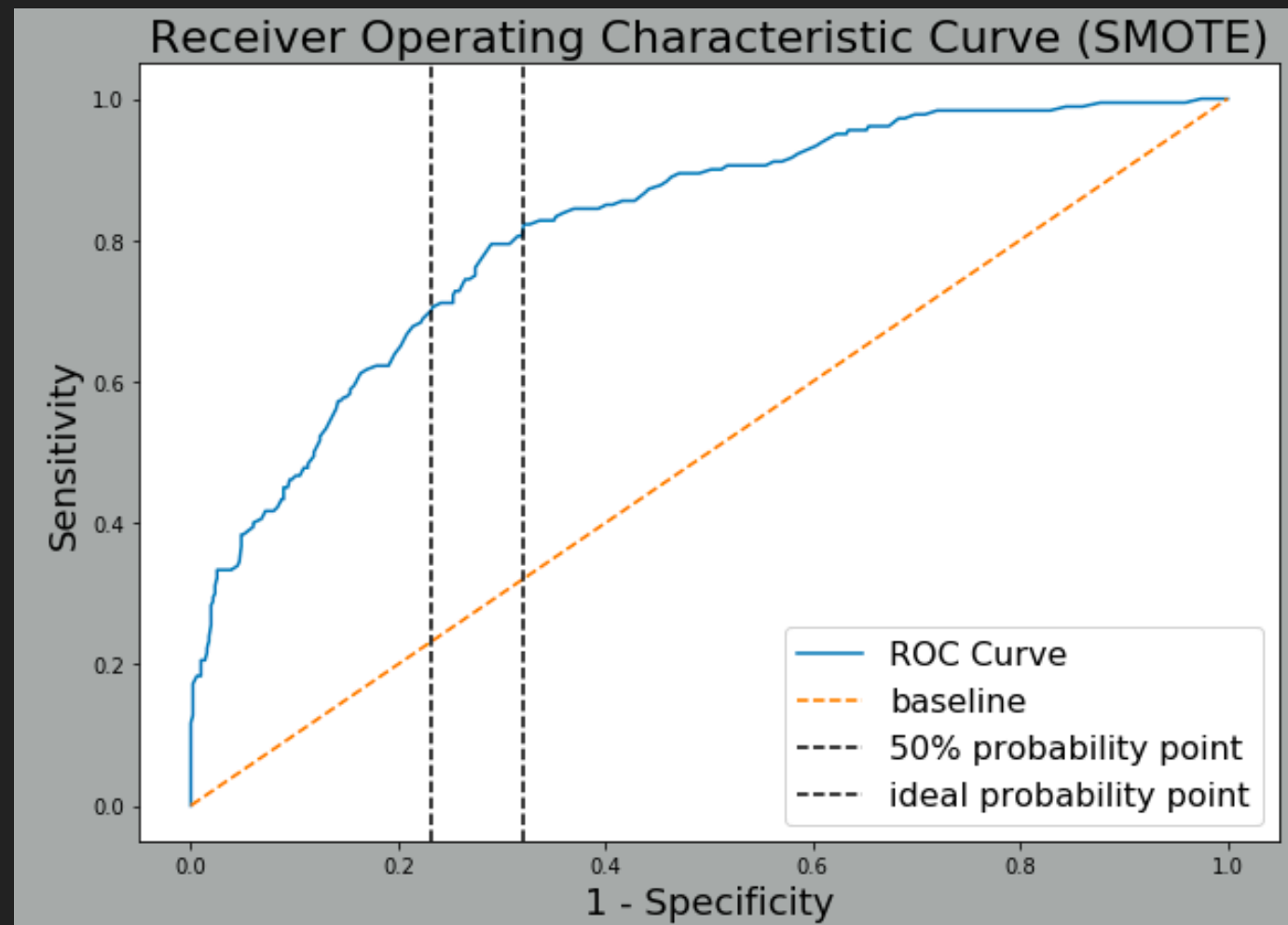
- | | |
|-------------|-------------|
| 1. ubi | 6. job |
| 2. welfare | 7. program |
| 3. debate | 8. business |
| 4. solution | 9. could |
| 5. freedom | 10. polling |

- ▶ Top 5 words/phrases least associated with Yang supporters

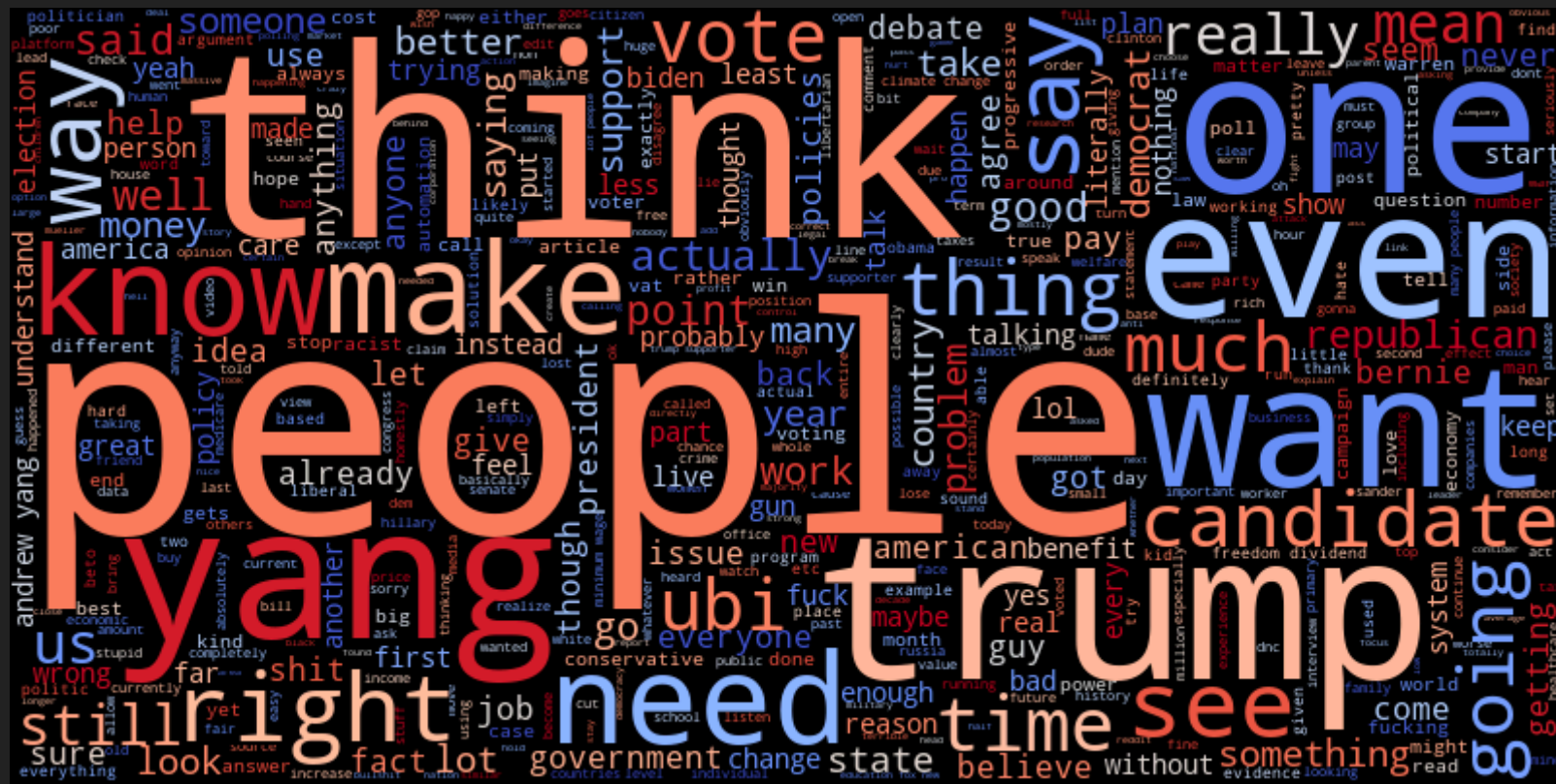
1. f***
2. democratic
3. house
4. hillary
5. corporate

MODEL RESULTS – ALL FEATURES

- Combining all my features and applying SMOTE, my sensitivity score was .6944, worse than my NLP model but applicable to all users



BONUS: TOPIC MODELING



BONUS: TOPIC MODELING

- ▶ Thankfully, there's a more useful model called Latent Dirichlet Allocation or LDA
- ▶ LDA was able to split out the topics that Yang supporters discussed on r/politics into 4 areas
 1. Trump / Republicans
 2. Debates / other 2020 candidates
 3. Automation / UBI
 4. Debt related

BONUS: TOPIC MODELING (DEMO)

CONCLUSION

- ▶ I was able to learn a number of things regarding Andrew Yang supporters
- ▶ Many of the interests I identified were obvious / known previously but some of the unexpected interests were crypto-currency, hip-hop fans, and marijuana enthusiasts
- ▶ Hopefully, if current Yang supporters with these interests share their support of Andrew Yang with others who have these interests, they have a receptive audience and a better success/conversion rate