

# Classifying Subreddits

## Conspiracy vs News

Presented by Nikhil Lonberg

# Computational propaganda upends American life

- “Computational propaganda” is the use of algorithms, automation, and big data to shape public life.
- 2019 Oxford University study finds social media misinformation campaigns contributing to global political instability and rise of authoritarianism.
- In America, hate speech, fake news, and conspiracy theories are commonplace on social media

# Timeline of Reddit content moderation

- In September, 2018, Reddit banned a QAnon conspiracy subreddit.
- In June, 2020, Reddit updated their hate speech policy to prohibit content that "promote[s] hate based on identity or vulnerability".
- New policy implemented immediately and 2000 subreddits, including r/The\_Donald, are removed
- Many users of banned subreddits migrated to different ones, creating a new set of controversial communities

# Project Goals and Questions

- Build binary classifier that distinguishes posts from r/news and r/conspiracy
- Determine what features differentiate news and conspiracy posts
- Optimize model for accuracy and sensitivity
- Propose method for conspiracy theory content moderation

# Process

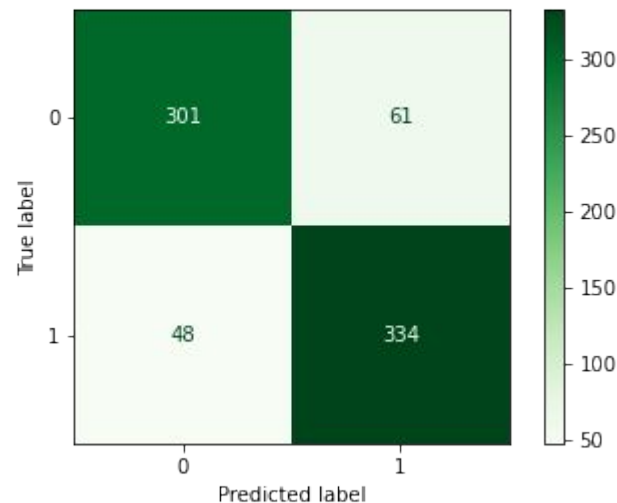
1. Scrape Reddit.com for 1500 r/news posts and 1500 r/conspiracy posts
2. Build a Bag of Words with CountVectorizer
3. Optimize models for accuracy and sensitivity
4. Propose method for flagging and moderating conspiracy theories

**trump (+135)**  
**just (+89)**  
**conspiracy (+76)**  
**biden (+57)**  
**people (+57)**  
**like (+51)**  
**think (+50)**  
**qanon (+42)**  
**media (+41)**  
**know (+37)**

**police (-122)**  
**man (-72)**  
**says (-66)**  
**coronavirus (-64)**  
**california (-57)**  
**portland (-38)**  
**arrested (-34)**  
**students (-33)**  
**mail (-31)**  
**postal (-29)**

Data Scrape → Bag of Words → **Modeling** → Proposal

- **Proposed Models:** Multinomial Naive Bayes (MNB), Gaussian Naive Bayes (GNB), and Logistic Regression
- **Accuracy Rank:** 1. MNB      2. Logistic Regression      3. GNB
- **Sensitivity Rank:** 1. MNB      2. Logistic Regression      3. GNB
- **Final Model:** Multinomial Naive Bayes
  - Accuracy: 87%
  - Sensitivity: 92%



**Data Scrape** —————> **Bag of Words** —————> **Modeling** —————> **Proposal**

- Use Multinomial Naive Bayes Classifier with high sensitivity hyperparameters to flag suspicious posts
- Have human moderator review flagged posts to determine if they are conspiracy theories
- Regularly retrain Naive Bayes Classifier on latest headlines