Tailored Audiences on Twitter

Insights from ~1,000,000 tweets about the 2016 Democratic National Convention

Jake Mullins

About

- US Navy Veteran and Reservist, Hospital Corpsman

- Studied chemistry at Georgia Tech 2014-2018

- Deployed to Cuba in 2019, Leading Petty Officer of Logistics for the Joint Medical Group

Motivation

Previous Project:

- 100,000 DNC tweets.

- Performed Network Analysis

- Classified users under 5 distinct communities

More Data:

- Collect ~900,000 tweets

- Leverage AWS Sagemaker for more computational power.

Refine Models:

- Use clustering methods in addition to network analysis.

- Identify more groups and elicit detailed information about their demographics.
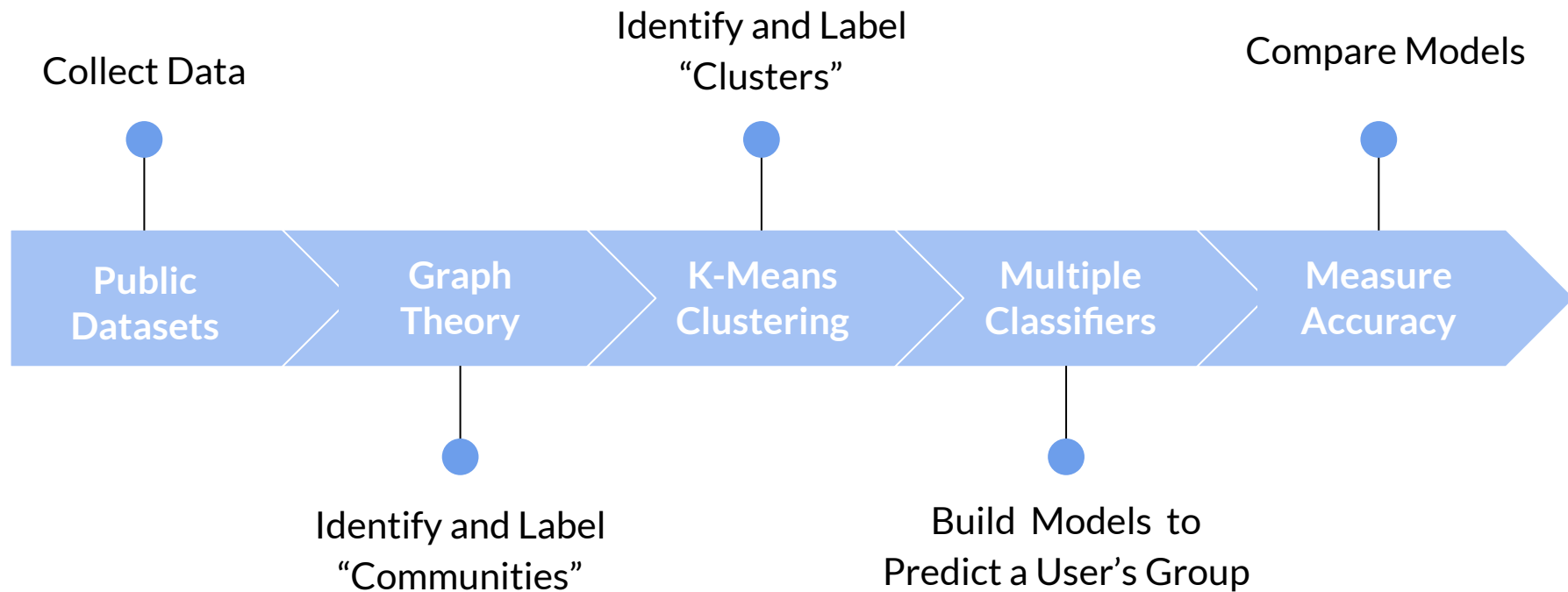
Background - Targeted Twitter Ads

Standard Audience Targeting:

- Age

- Gender

- Geography

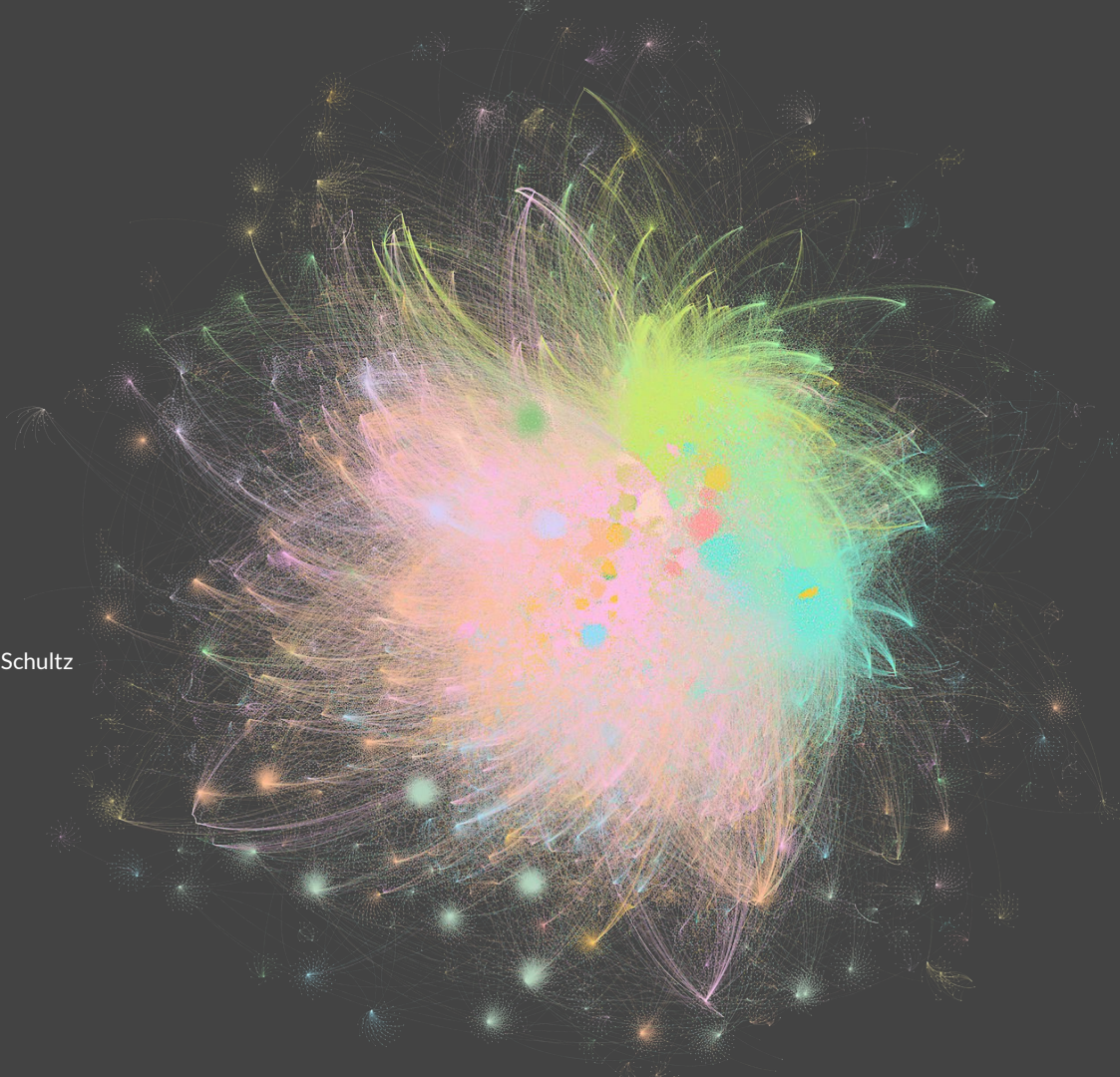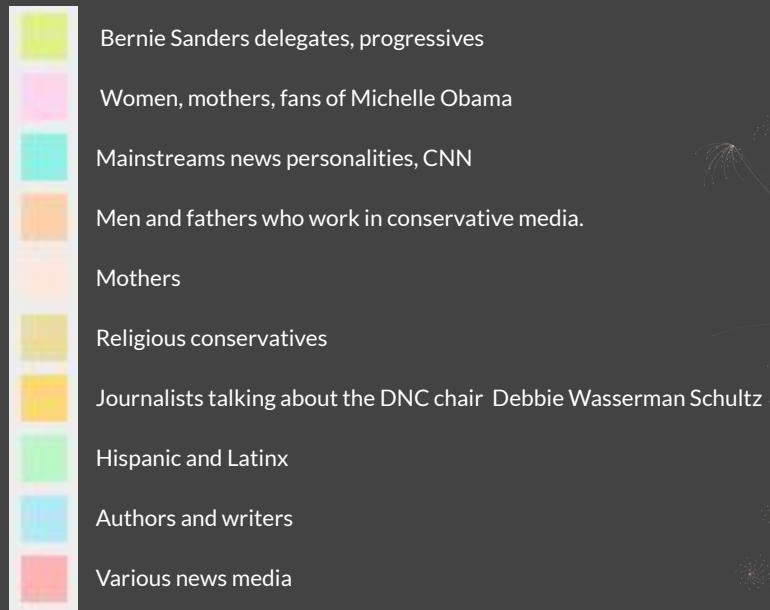- Generic interests (eg "shopping", "sports")

Tailored Audiences:

- Mothers who are fans of Michelle Obama

- Men who are fathers and religious

- Teenagers who support Bernie Sanders

- And more...

Collect Data

Identify and Label
"Clusters"

Compare Models

Public Datasets → **Graph Theory** → **K-Means Clustering** → **Multiple Classifiers** → **Measure Accuracy**

Identify and Label
"Communities"

Build Models to
Predict a User's Group

Identifying and Labeling Communities:

1. Create graph objects from tweet metadata .
   a. Nodes = Twitter Users
   b. Edges = Directed Tweets (@mentions, replies)
2. Use Gephi software to group users into communities.
3. Summarize the text from each community and extract topics.

Groups users based on *who* they talk to

Bernie Sanders delegates, progressives

Women, mothers, fans of Michelle Obama

Mainstreams news personalities, CNN

Men and fathers who work in conservative media.

Mothers

Religious conservatives

Journalists talking about the DNC chair Debbie Wasserman Schultz

Hispanic and Latinx

Authors and writers

Various news media

Identifying and Labeling Clusters

1. Convert tweet text into vectors (GloVe model)
2. Group users into clusters based on similarity of their language (k-means clustering)
3. Summarize the text from each community and extract topics

Groups users based on *what* they talk about

# Clusters vs. Communities

Cluster Labels

- DNC chair Debbie Wasserman Schultz

- Women's issues

- Expressing concern about trump and nuclear codes

- Love, God, good vibes

- Watching the debate with friends, supporting Hillary Clinton

- American Pride

- Conservative sports fans

- Barack and Michelle Obama

- Barack and Michelle Obama

- Support for Bernie Sanders (mostly young users)

Community Labels

- Bernie Sanders delegates, progressives

- Women, mothers, fans of Michelle Obama

- Mainstreams news personalities, CNN

- Men and fathers who work in conservative media.

- Mothers

- Religious conservatives

- Journalists talking about the DNC chair Debbie Wasserman Schultz

- Hispanic and Latinx

- Authors and writers

- Various news media
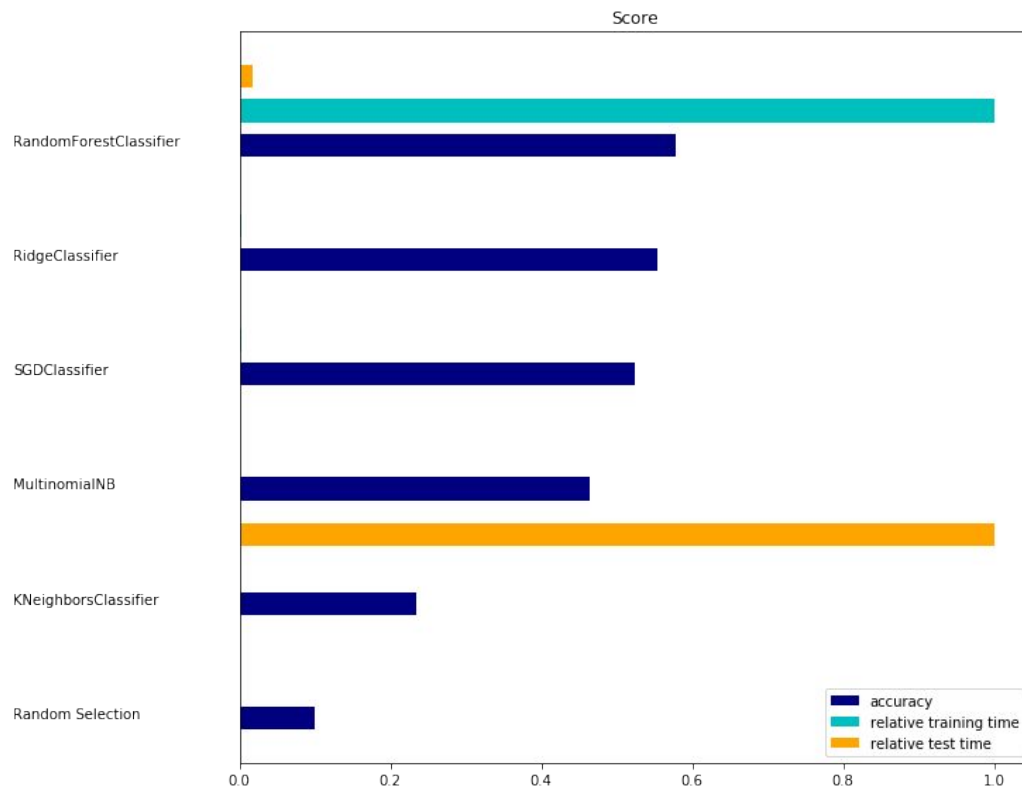
Community Classifier Benchmarks:

- Accuracy of randomly selecting desired group: **~10%**

- Accuracy of my best model: **55.76%**
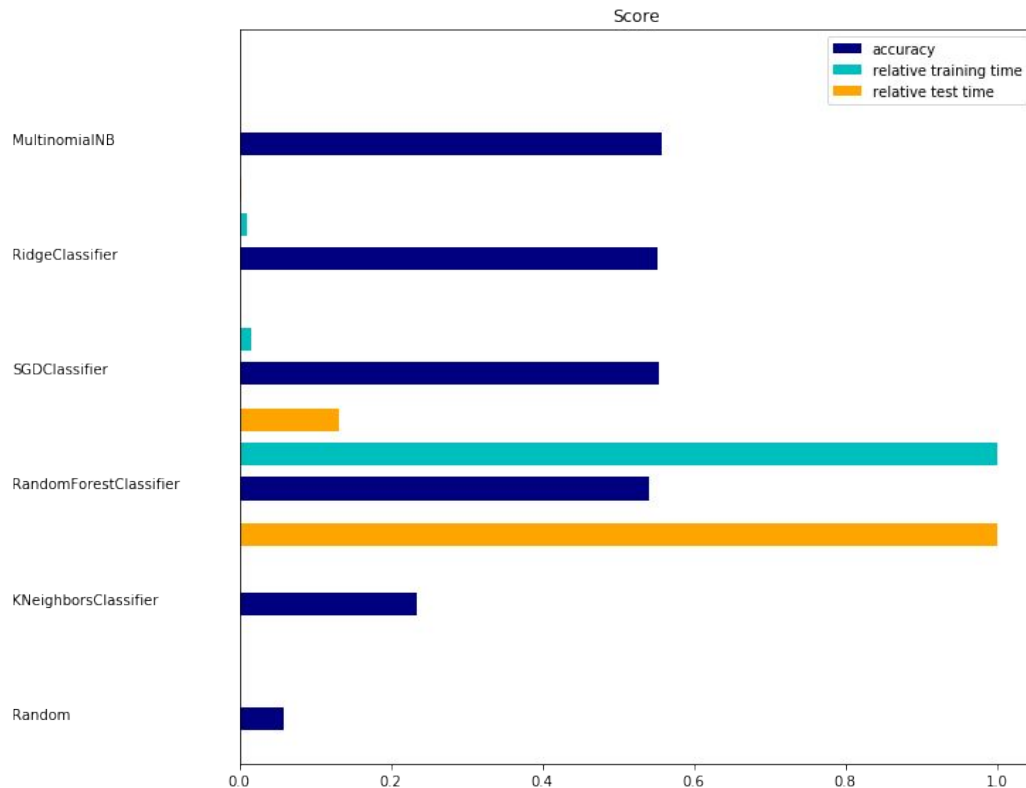
Cluster Classifier Benchmarks:

- Accuracy of randomly selecting desired group: **~10%**

- Accuracy of my best model: **55.82%**

Other Potential Use Cases:

Customer Segmentation

Use grouping data to direct ad campaigns tailored toward existing customers of various demographics

Public Health

Use tweets about vaccinations, covid-19 etc. to identify clusters of disinformation.

Questions:

Contact:

- LinkedIn.com/in/jakecmullins

- Github.com /jakemull13

- Jakemull13@gmail.com