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.

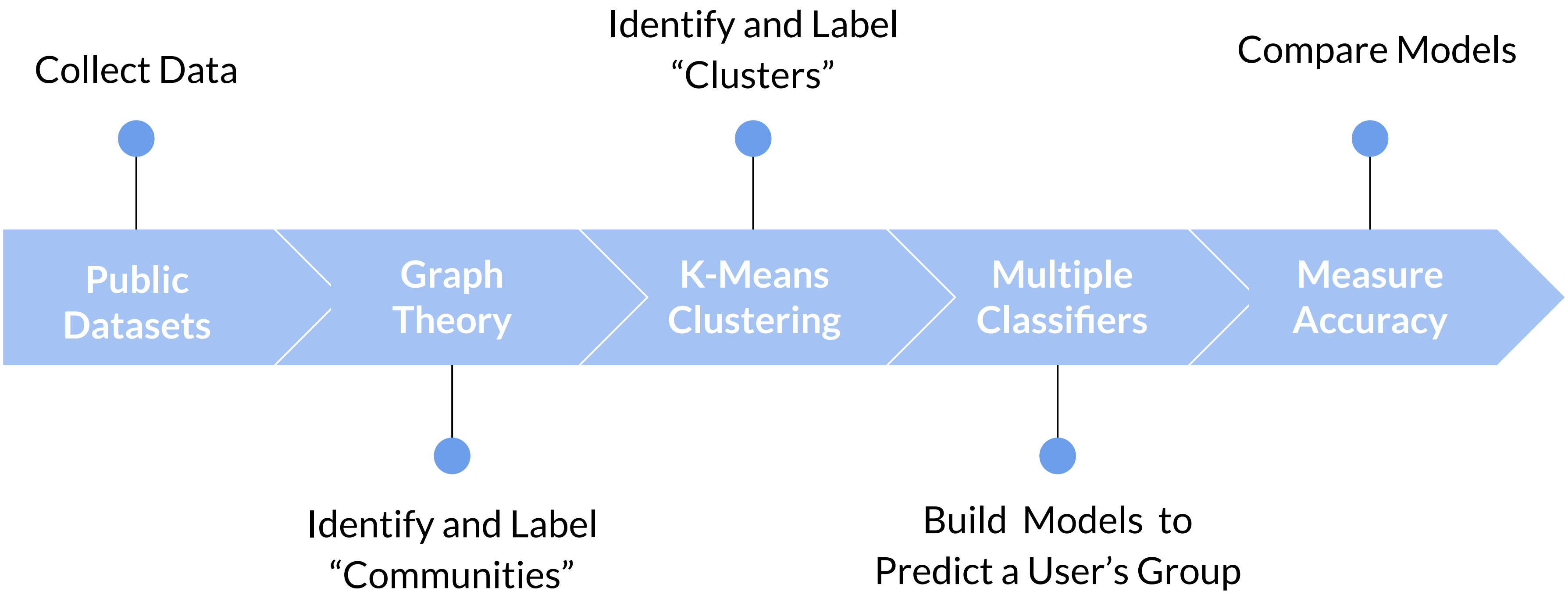
Background - Targeted Twitter Ads

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

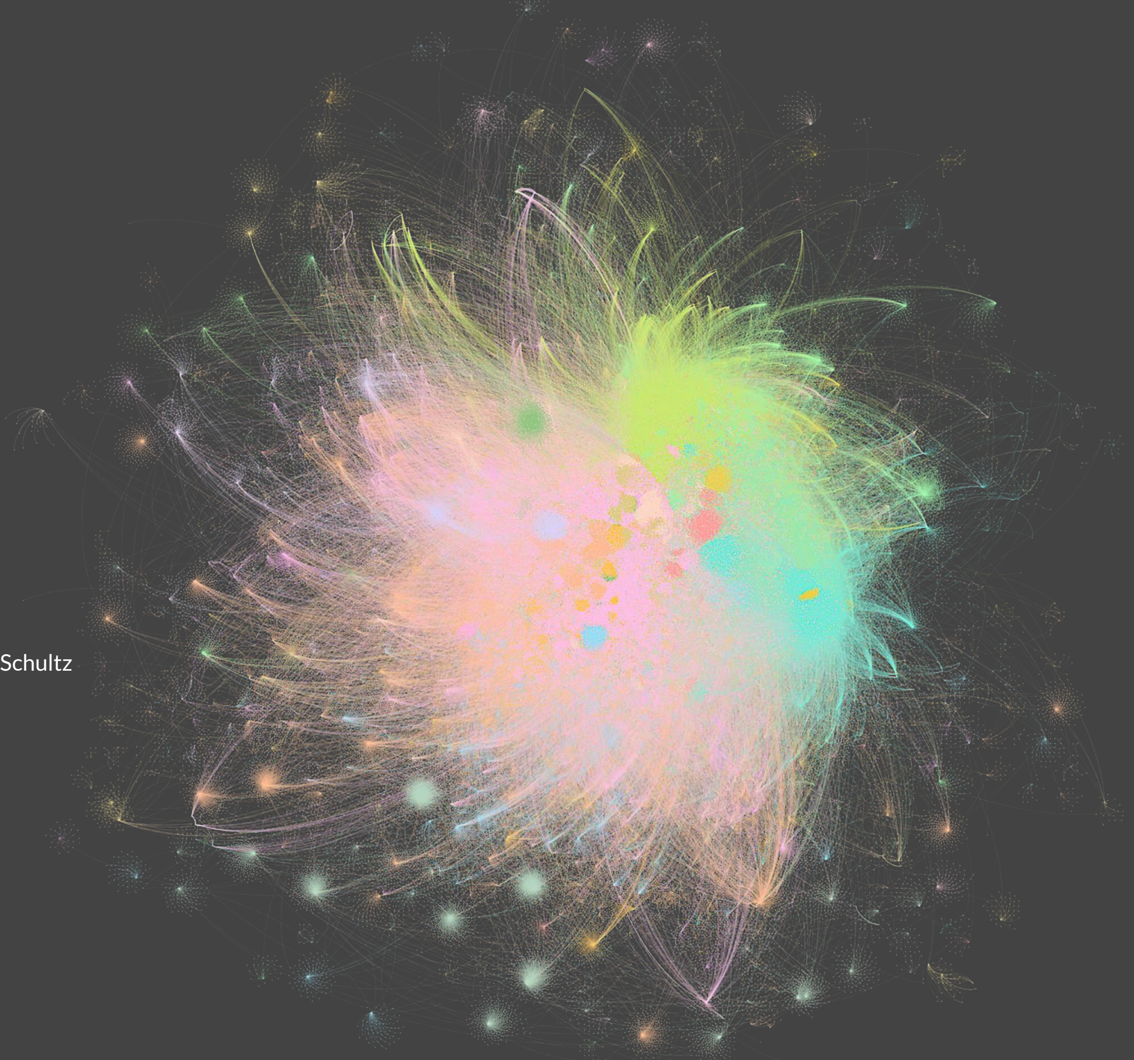
Religious conservatives

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

Clusters vs. Communities

Cluster Labels

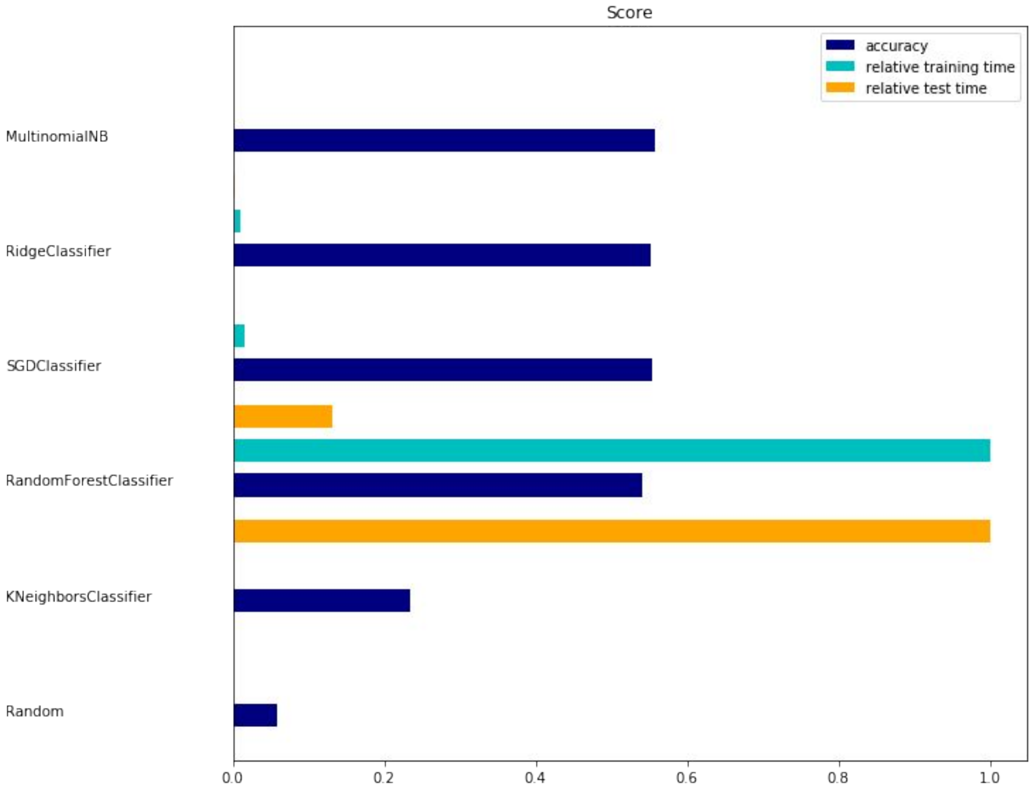
- DNC chair Debbie Wasserman Schultz
- Women's issues
- Expressing concern about trump and nuclear codes
- Love, God, good vibes
- Watching the debate with friends, supporting Hillary Clinton
- American Pride
- Conservative sports fans
- Barack and Michelle Obama (positively)
- Barack and Michelle Obama (positively)
- Support for Bernie Sanders (mostly young users)

Community Labels

- Bernie Sanders supporters, 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

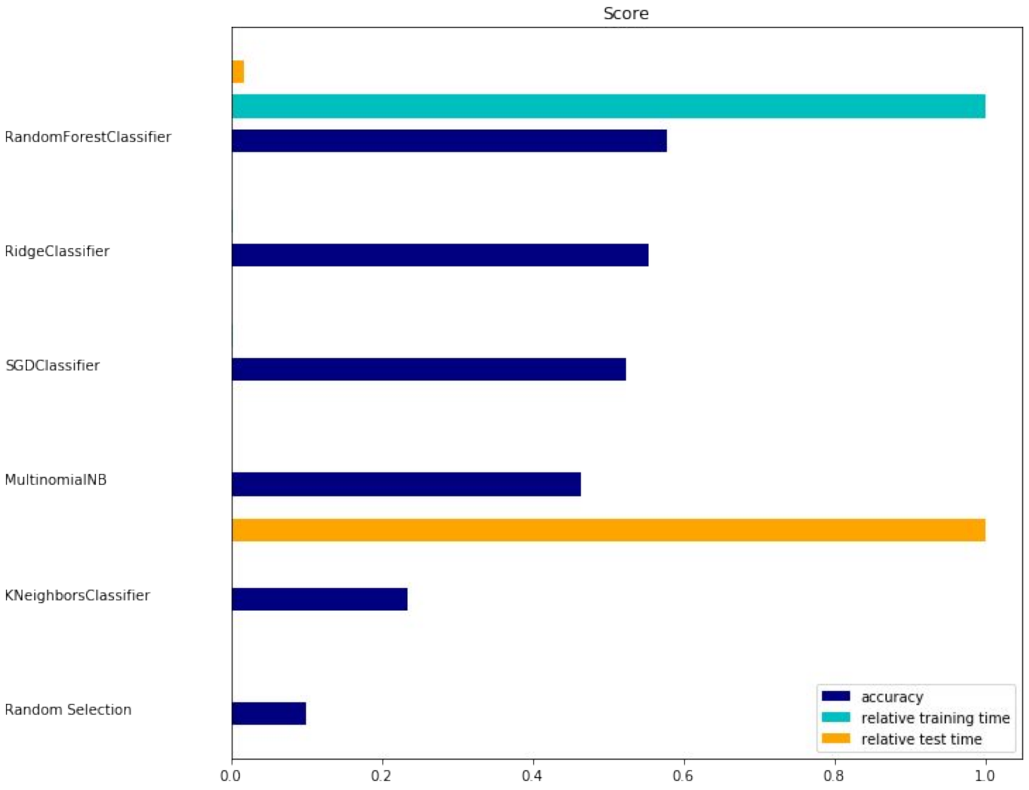
Community Classifier Benchmarks:

- 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

Acknowledgements:

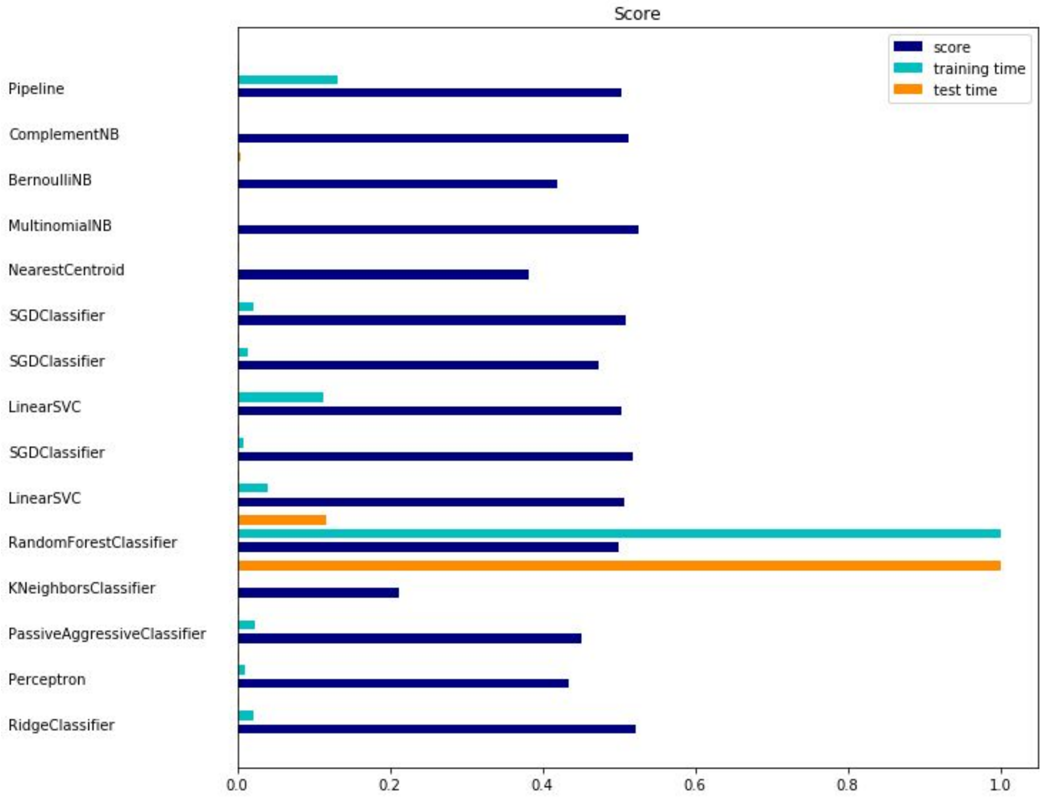
1. Galvanize Instructional Team
2. Sed do eiusmod tempor incididunt ut labore
3. Sed do eiusmod tempor incididunt ut labore

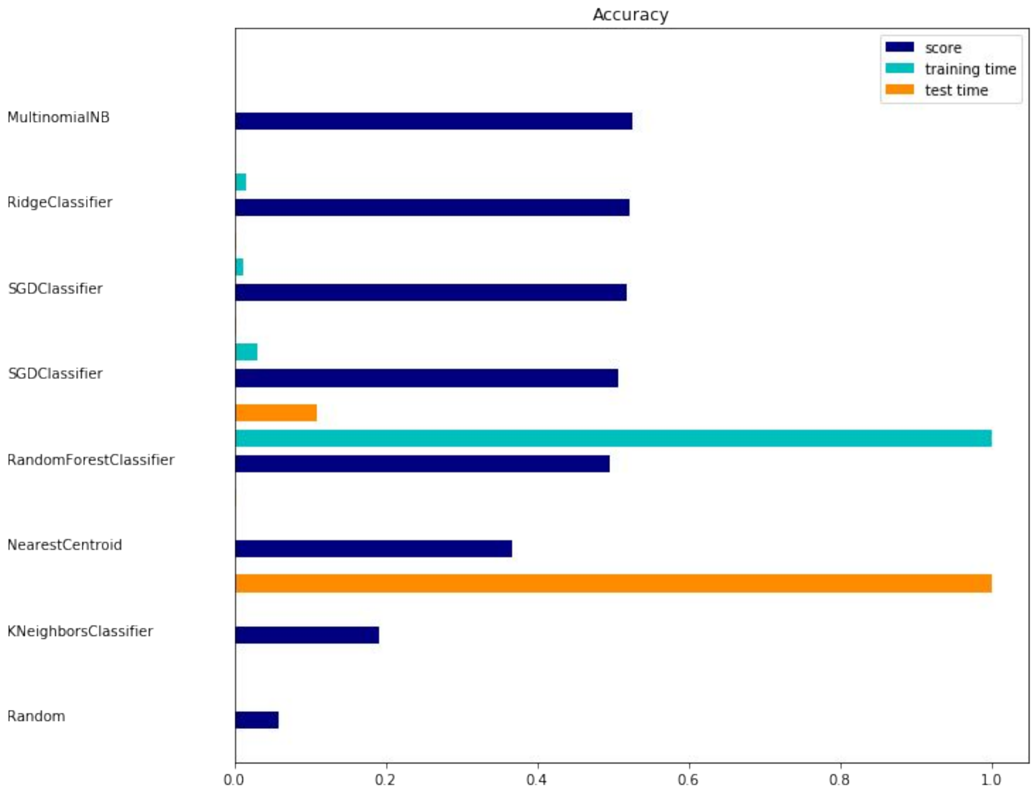
Contact:

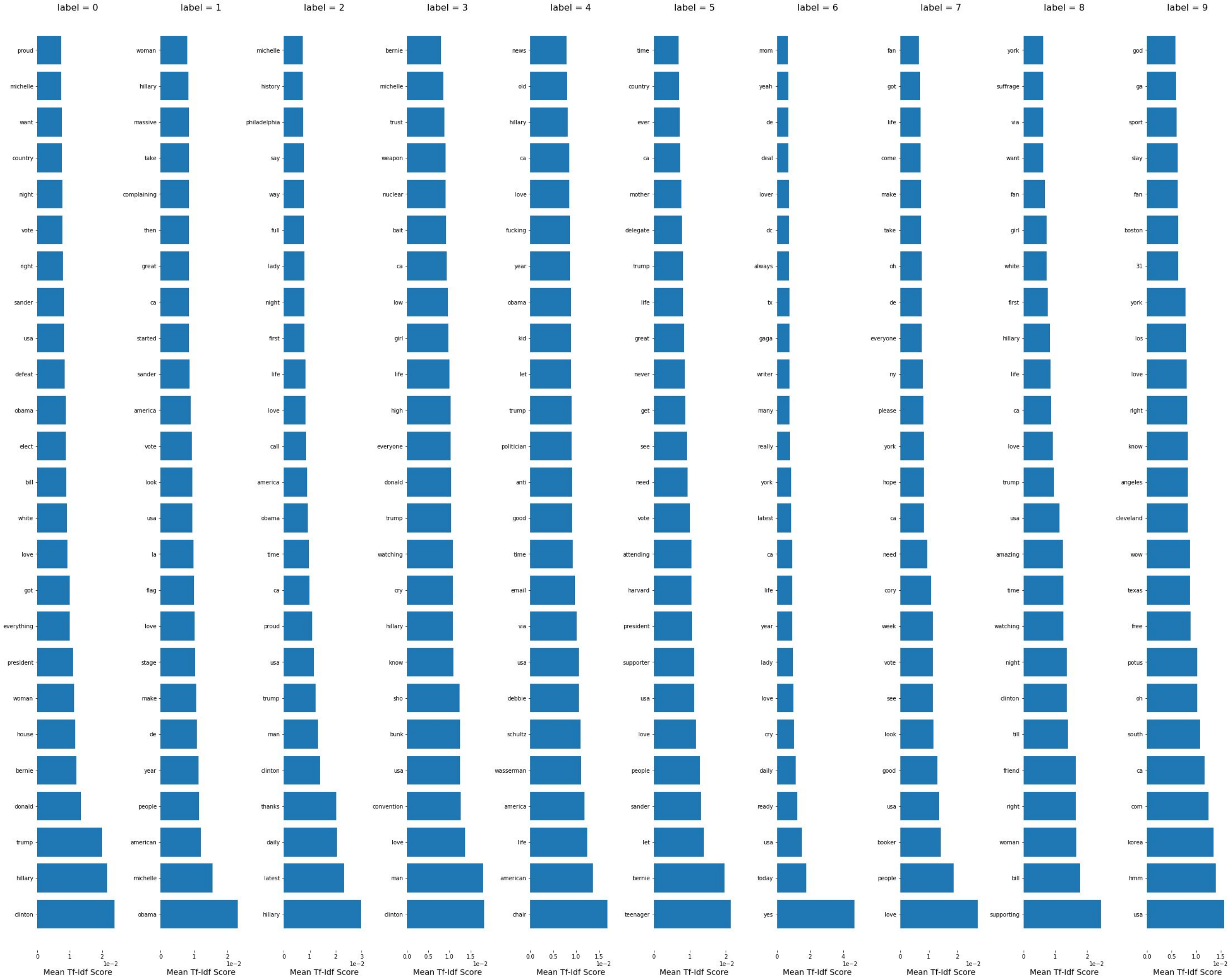
- LinkedIn.com/in/jakecmullins
- Github.com /jakemull13
- Jakemull13@gmail.com

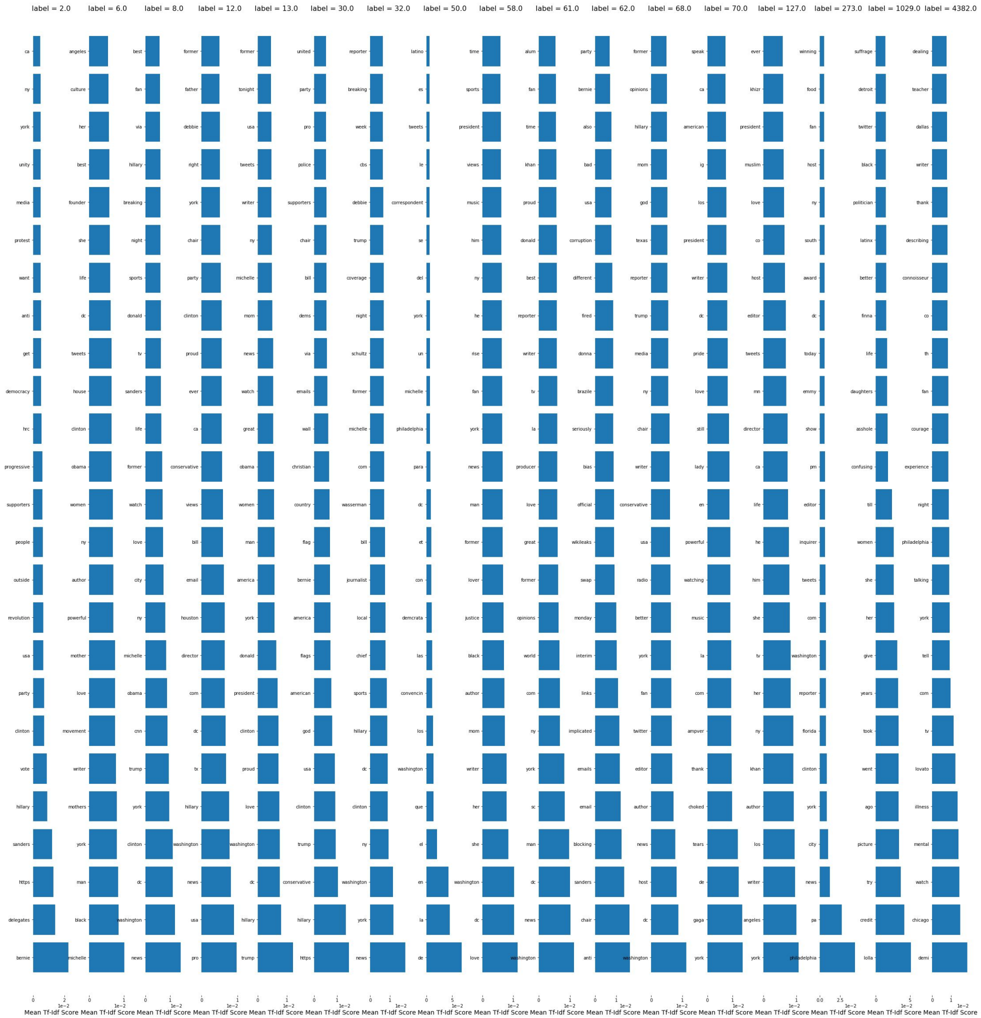
Additional Plots:

30	(17.66%)
13	(15.31%)
2	(9.77%)
6	(5.53%)
8	(5.01%)
61	(4.75%)
58	(3.24%)
68	(3.05%)
127	(2.58%)
12	(2.43%)
70	(2.37%)
1029	(2.11%)
62	(1.5%)
50	(1.41%)
273	(1.39%)
4382	(1.25%)
32	(1.19%)


Score
score
training time
test time
Pipeline
ComplementNB
BernoulliNB
MultinomialNB
NearestCentroid
SGDClassifier
SGDClassifier
LinearSVC
SGDClassifier
LinearSVC
RandomForestClassifier
KNeighborsClassifier
PassiveAggressiveClassifier
Perceptron
RidgeClassifier
0.0
0.2
0.4
0.6
0.8
1.0


Accuracy
score
training time
test time
MultinomialNB
RidgeClassifier
SGDClassifier
SGDClassifier
RandomForestClassifier
NearestCentroid
KNeighborsClassifier
Random
0.0
0.2
0.4
0.6
0.8
1.0


label = 0
proud
michelle
want
country
night
vote
right
sander
usa
defeat
obama
elect
bill
white
love
got
everything
president
woman
house
bernie
donald
trump
hillary
clinton
0 1 2
1e−2
Mean Tf-Idf Score
label = 1
woman
hillary
massive
take
complaining
then
great
ca
started
sander
america
vote
look
usa
la
flag
love
stage
make
de
year
people
american
michelle
obama
0 1 2
1e−2
Mean Tf-Idf Score
label = 2
michelle
history
philadelphia
say
way
full
lady
night
first
life
love
call
america
obama
time
ca
proud
usa
trump
man
clinton
thanks
daily
latest
hillary
0 1 2 3
1e−2
Mean Tf-Idf Score
label = 3
bernie
michelle
trust
weapon
nuclear
bait
ca
low
girl
life
high
everyone
donald
trump
watching
cry
hillary
know
sho
bunk
usa
convention
love
man
clinton
0.0 0.5 1.0 1.5
1e−2
Mean Tf-Idf Score
label = 4
news
old
hillary
ca
love
fucking
year
obama
kid
let
trump
politician
anti
good
time
email
via
usa
debbie
schultz
wasserman
america
life
american
chair
0.0 0.5 1.0 1.5
1e−2
Mean Tf-Idf Score
label = 5
time
country
ever
ca
mother
delegate
trump
life
great
never
get
see
need
vote
attending
harvard
president
supporter
usa
love
people
sander
let
bernie
teenager
0 1 2
1e−2
Mean Tf-Idf Score
label = 6
mom
yeah
de
deal
lover
dc
always
tx
gaga
writer
many
really
york
latest
ca
life
year
lady
love
cry
daily
ready
usa
today
yes
0 2 4
1e−2
Mean Tf-Idf Score
label = 7
fan
got
life
come
make
take
oh
de
everyone
ny
please
york
hope
ca
need
cory
week
vote
see
look
good
usa
booker
people
love
0 1 2
1e−2
Mean Tf-Idf Score
label = 8
york
suffrage
via
want
fan
girl
white
first
hillary
life
ca
love
trump
usa
amazing
time
watching
night
clinton
till
friend
right
woman
bill
supporting
0 1 2
1e−2
Mean Tf-Idf Score
label = 9
god
ga
sport
slay
fan
boston
31
york
los
love
right
know
angeles
cleveland
wow
texas
free
potus
oh
south
ca
com
korea
hmm
usa
0.0 0.5 1.0 1.5
1e−2
Mean Tf-Idf Score


label = 2.0
ca, ny, york, unity, media, protest, want, anti, get, democracy, hrc, progressive, supporters, people, outside, revolution, usa, party, clinton, vote, hillary, sanders, https, delegates, bernie
label = 6.0
angeles, culture, her, best, founder, she, life, dc, tweets, house, clinton, obama, women, ny, author, powerful, mother, love, movement, writer, mothers, york, man, black, michelle
label = 8.0
best, fan, via, hillary, breaking, night, sports, donald, tv, sanders, life, former, watch, love, city, ny, michelle, obama, cnn, trump, york, clinton, dc, washington, news
label = 12.0
former, father, debbie, right, york, chair, party, clinton, proud, ever, ca, conservative, views, bill, email, houston, director, com, dc, tx, hillary, washington, news, usa, pro
label = 13.0
former, tonight, usa, tweets, writer, ny, michelle, mom, news, watch, great, obama, women, man, america, york, donald, president, clinton, proud, love, washington, dc, hillary, trump
label = 30.0
united, party, pro, police, supporters, chair, bill, dems, via, emails, wall, christian, country, flag, bernie, america, flags, american, god, usa, clinton, trump, conservative, hillary, https
label = 32.0
reporter, breaking, week, cbs, debbie, trump, coverage, night, schultz, former, michelle, com, wasserman, bill, journalist, local, chief, sports, hillary, dc, clinton, ny, washington, york, news
label = 50.0
latino, es, tweets, le, correspondent, se, del, york, un, michelle, philadelphia, para, dc, et, con, demcrata, las, convencin, los, washington, que, el, en, la, de
label = 58.0
time, sports, president, views, music, him, ny, he, rise, fan, york, news, man, former, lover, justice, black, author, mom, writer, her, she, washington, dc, love
label = 61.0
alum, fan, time, khan, proud, donald, best, reporter, writer, tv, la, producer, love, great, former, opinions, world, com, ny, york, sc, man, dc, news, washington
label = 62.0
party, bernie, also, bad, usa, corruption, different, fired, donna, brazile, seriously, bias, official, wikileaks, swap, monday, interim, links, implicated, emails, email, blocking, sanders, chair, anti
label = 68.0
former, opinions, hillary, mom, god, texas, reporter, trump, media, ny, chair, writer, conservative, usa, radio, better, york, fan, twitter, editor, author, news, host, dc, washington
label = 70.0
speak, ca, american, ig, los, president, writer, dc, pride, love, still, lady, en, powerful, watching, music, la, com, ampver, thank, choked, tears, de, gaga, york
label = 127.0
ever, khizr, president, muslim, love, co, host, editor, tweets, mn, director, ca, life, he, him, she, tv, her, ny, khan, author, los, writer, angeles, york
label = 273.0
winning, food, fan, host, ny, south, award, dc, today, emmy, show, pm, editor, inquirer, tweets, com, washington, reporter, florida, clinton, york, city, news, pa, philadelphia
label = 1029.0
suffrage, detroit, twitter, black, politician, latinx, better, finna, life, daughters, asshole, confusing, till, women, she, her, give, years, took, went, ago, picture, try, credit, lolla
label = 4382.0
dealing, teacher, dallas, writer, thank, describing, connoisseur, co, th, fan, courage, experience, night, philadelphia, talking, york, tell, com, tv, lovato, illness, mental, watch, chicago, demi
Mean Tf-Idf Score