'Covid Vaccine' Tweets

Sentiment Analysis with Natural Language Processing

by David Bruce

Content Outline

01 Context

02 Preview Results

03 Data & Methods

04 Recommendations

05 Next Steps

A Global Pandemic

January 20, 2020

First known case of the novel Coronavirus arrives in the US

September 22, 2020

US surpasses 200,000 deaths

December 14, 2020

First inoculation given to public of FDA approved Pfizer vaccine

Results Preview

Multiclass Classification

Model	Train F1	Test F1
Naive Bayes	0.723	0.709
SVM	0.829	0.815
RandomForest	0.312	0.311

CRISP-DM

An Iterative Process

Webscraping Twitter

Twint

Data Cleaning

Pandas

NLP & EDA

NLTK, SpaCy, VaderSentiment

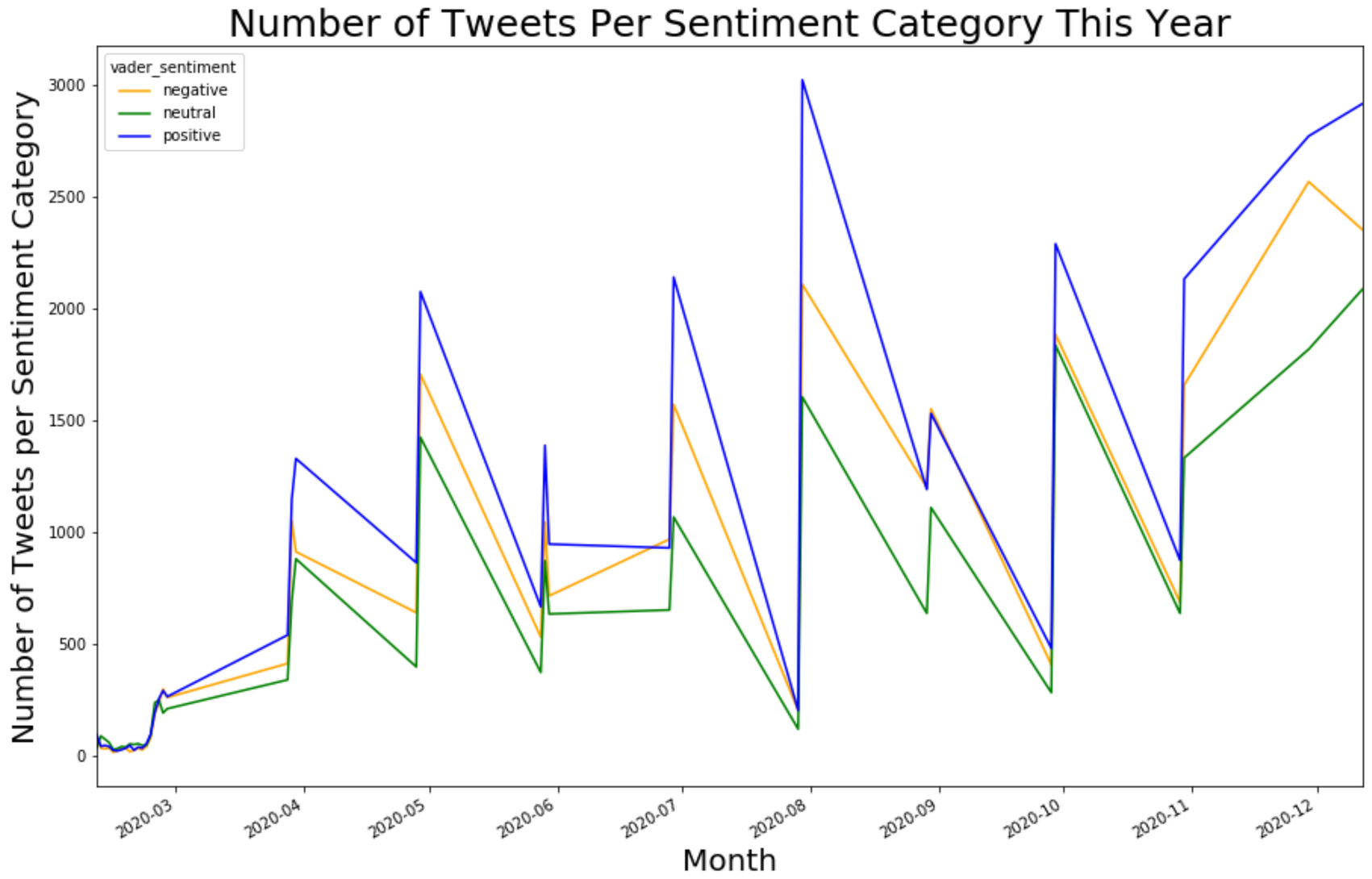
Modeling

Scikit Learn

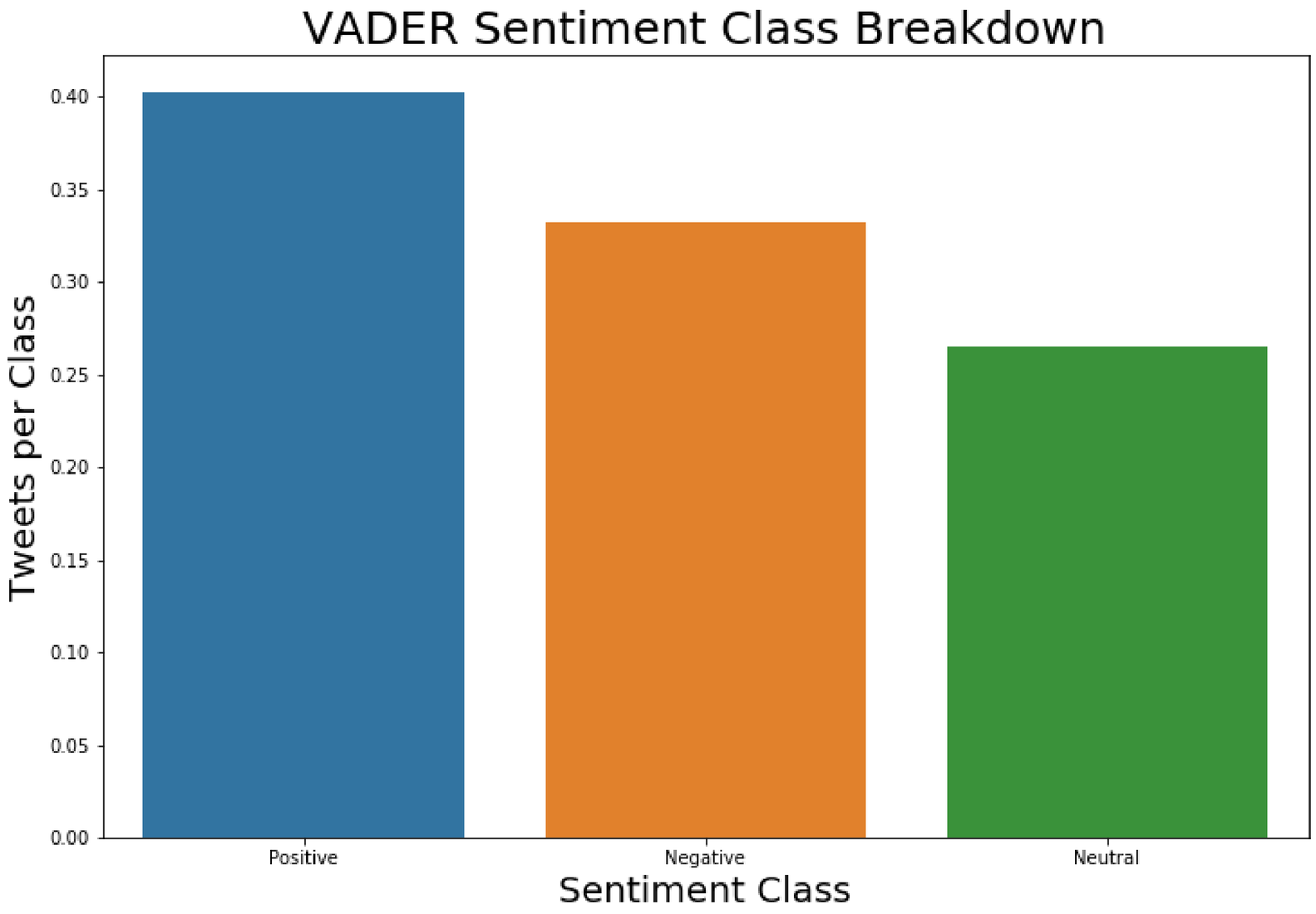
Reiterate

Also used: Matplotlib, Seaborn, WordCloud, and Numpy

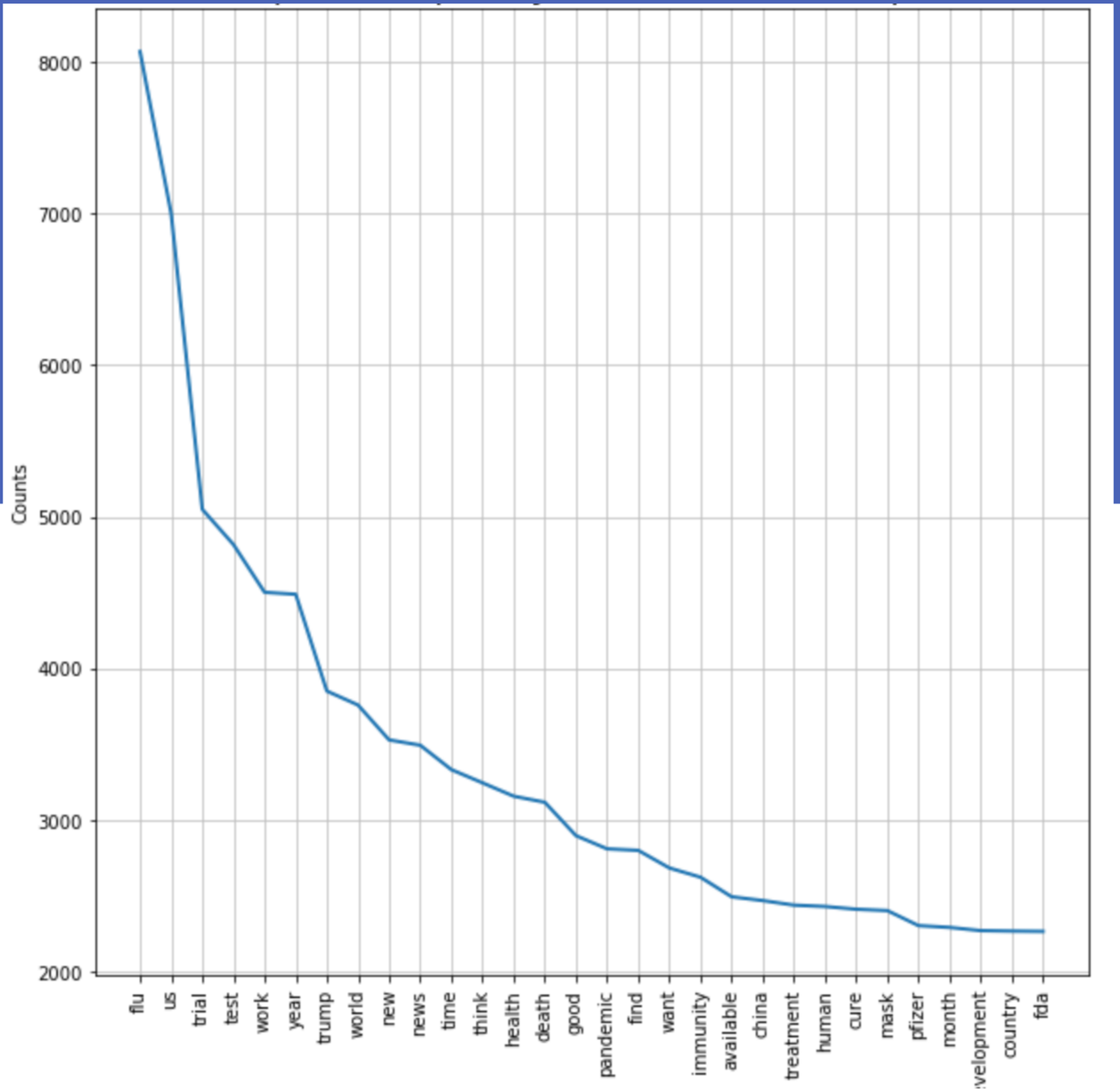
DATA ANALYSIS - 1



DATA ANALYSIS - 2

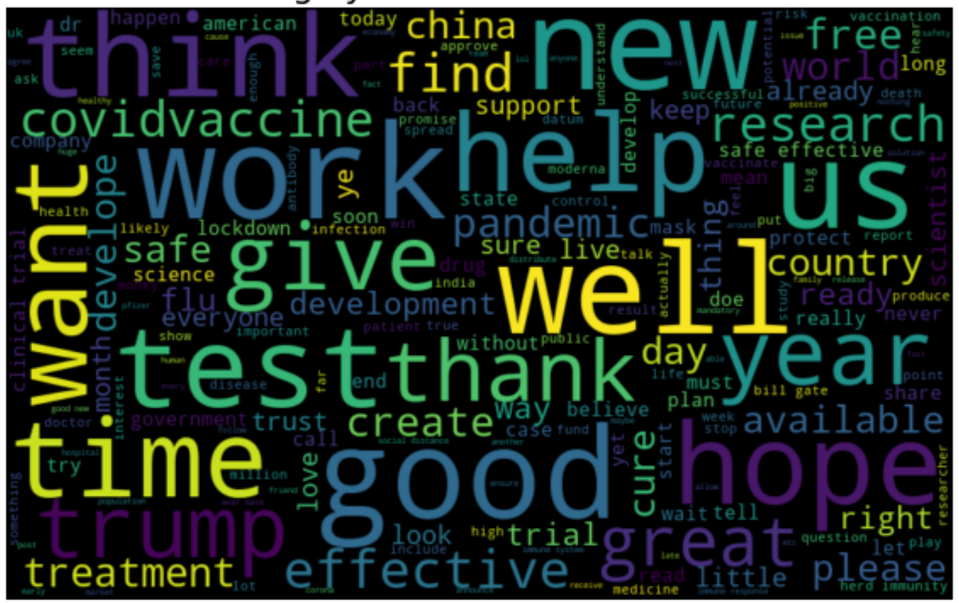


DATA ANALYSIS - 3

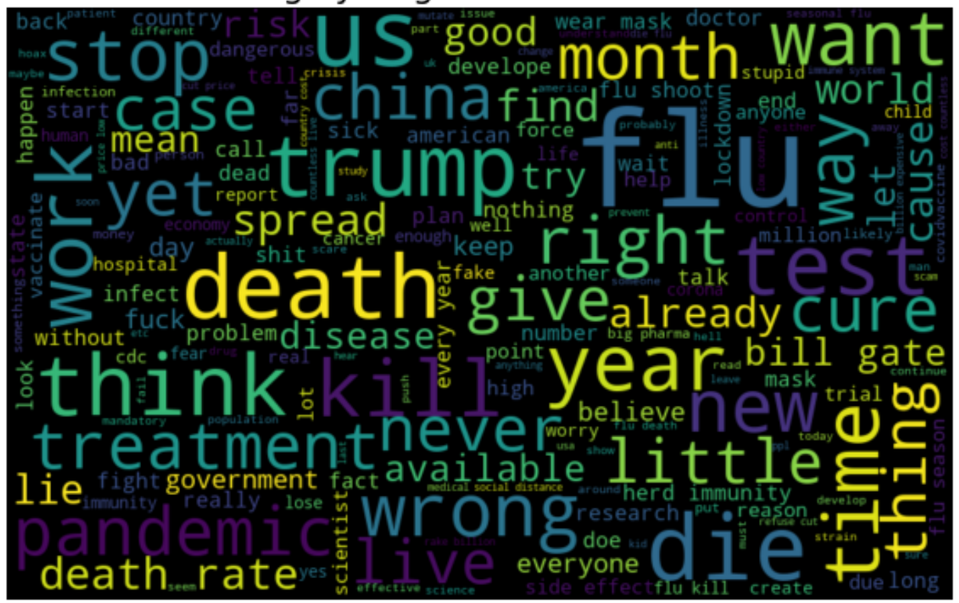


DATA ANALYSIS - 4

Highly Positive Wordcloud



Highly Negative Wordcloud


us
stop
flu
trump
want
month
good
wear mask
doctor
case
china
find
work
yet
right
test
death
give
already
cure
think
kill
year
new
bill gate
treatment
never
available
little
lie
wrong
pandemic
live
die
time
thing
rate
disease
risk
dangerous
develop
stupid
world
way
cause
mean
spread
nothing
keep
flu shoot
lockdown
anyone
end
hospital
infect
fake
another
talk
million
mask
trial
believe
herd immunity
reason
everyone
side effect
flu kill
create
long
back
country
different
hoax
maybe
happen
infection
start
human
bad
dead
report
sick
american
force
life
wait
try
help
economy
cancer
enough
shit
day
fuck
problem
without
cdc
fear
drug
real
lot
fight
government
fact
immunity
really
lose
yes
effective
science
vaccinate
state
something
look
every
year
high
worry
use
show
medical social distance
research
doe
kid
refuse cut
strain
due
flu season
covidvaccine
scam
hell
big pharma
point
anything
read
leave
continue
today
develop
must
put
around
make
part
understand
issue
crisis
price
tell
far
country cases
illness
probably
anti
prevent
control
let
child
stupid
lockdown
well
plan
nothing
keep
another
talk
already
bill gate
mask
trial
believe
herd immunity
reason
everyone
side effect
flu kill
create
long
due
flu season
thing
time
die
wrong
pandemic
live
death
rate
lie
scientist
treatment
never
available
little
think
kill
year
new
death
give
already
cure
test
right
flu
trump
want
month
good
wear mask
doctor
stop
us
case
china
find
work
yet
spread
mean
dangerous
develop
stupid
world
way
cause
flu shoot
lockdown
anyone
end
hospital
infect
fake
another
talk
million
mask
trial
believe
herd immunity
reason
everyone
side effect
flu kill
create
long
due
flu season
thing
time
die
wrong
pandemic
live
death
rate
lie
scientist
treatment
never
available
little
think
kill
year
new
bill gate

Next Steps

More Data

Location Based Tweets

LDA Topic Modeling with Gensim

Thank You.

Any Questions?

Email Address
david.bruce14@gmail.com

GitHub
@davidlebruce

Twitter
@dgbruce1