Detecting Fake Health News

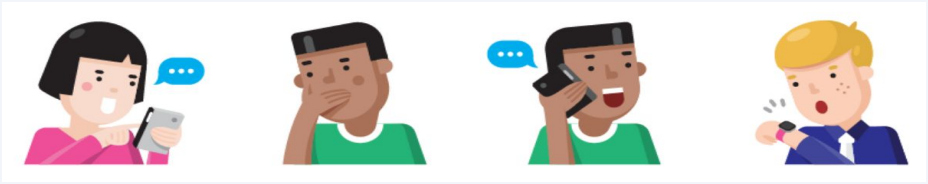
- Vaccines

By
Mohammed Rehan
Phil Greene
Sheethal Sridhar
Shubham Puri

Contents

- ❑ Current Scenario
- ❑ Objective
- ❑ Process
- ❑ Tools used
- ❑ Data gathering
- ❑ Data Description
- ❑ Fact Repository
- ❑ Classification
- ❑ Data Cleaning
- ❑ Modelling
- ❑ Analysis

Current Scenario

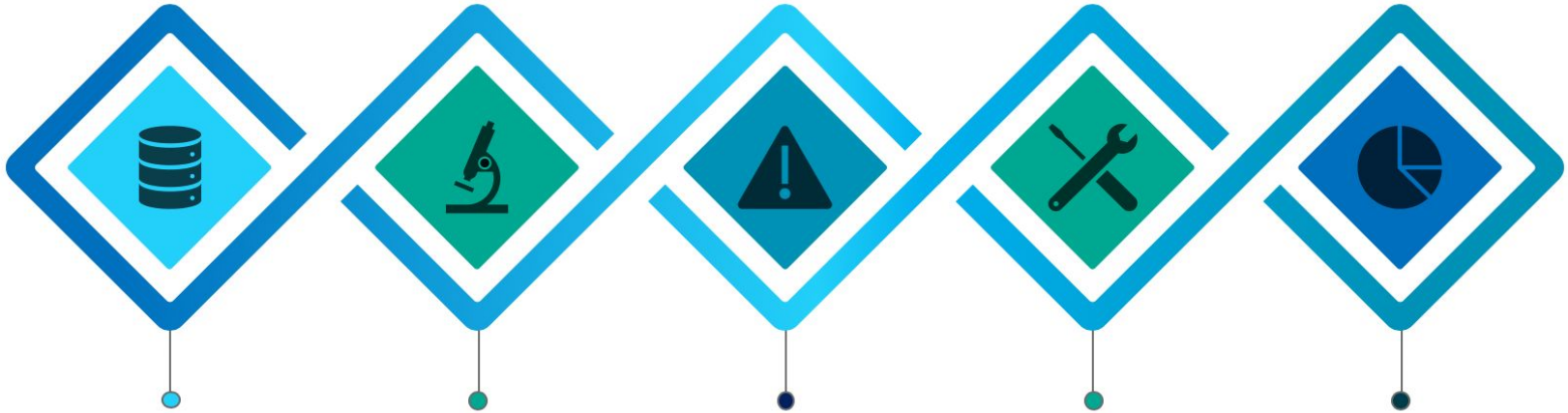


- False news reaches more people than the truth
- Motivations of its creators vary, but fake medical news can earn clicks and “likes,” which can translate to ad revenue, or further become an agenda for large-scale investments
- False information can also drive the sale of expensive, unproven treatments
- Unproven remedies may be harmful, waste money and sometimes, using these remedies keeps people from getting the medical treatment they need

Objective

- Detect fake tweets about ‘vaccines’ using Machine Learning models
- Analyzing the fake tweet characteristics

Process



Data Collection

- ❑ Collecting tweets relevant to vaccines
- ❑ Gathering a fact repository

Classification

- ❑ Manual Classification of tweets into buckets
- ❑ Classified claims into fake and not fake

Building a Model

- ❑ Using Google Auto ML
- ❑ Naïve Bayes Approach - Python

Testing

Using the model to classify the claims as fake or not

Analysis

Analysis of tweets and users characteristics

Tools Used

Data Gathering

- Used twitter API to collect tweets in R

- Collected tweets containing word vaccines and hashtag vaccines for the period of 4 weeks

(keywords: "vaccines", "#vaccines"," Vaccines", "#Vaccines")

- Gathered 20000 tweets, after removal of duplicates we were left with 18209 tweets

Data Description

Column	Type	Description
screen_name	char	The screen name, handle, or alias that this user identifies themselves with
text	char	The actual UTF-8 text of the status update
tweet_clean	char	tweets after cleaning
buckets	nominal	manually classified tweet types
fake	binary (True -1, False -0)	manually classified tweet is fake or not
pred_buckets	nominal	model output for type classification
pred_fake	binary	model output for fake classification
source	char	Utility used to post the Tweet, as an HTML-formatted string
is_quote	binary	Indicates whether this is a Quoted Tweet.
is_retweet	binary	Indicates whether this is a retweeted Tweet.
favorite_count	numeric	Indicates approximately how many times this Tweet has been liked by Twitter users
retweet_count	numeric	Number of times this Tweet has been retweeted
hashtags	char	Represents hashtags which have been parsed out of the Tweet text
media_type	char	Represents media elements uploaded with the Tweet
lang	char	language
location	char	The user-defined location for this account's profile. Not necessarily a location, nor machine-parseable
description	char	The user-defined UTF-8 string describing their account
followers_count	numeric	The number of followers this account currently has
friends_count	numeric	The number of users this account is following (AKA their "followings")
listed_count	numeric	The number of public lists that this user is a member of
statuses_count	numeric	The number of Tweets (including retweets) issued by the user.
favourites_count	numeric	The number of Tweets this user has liked in the account's lifetime
verified	binary	When true, indicates that the user has a verified account

Facts Repository

Analyzed multiple sources to gather a fact repository of vaccines

There were 16 most prominent myths among all the sources referred regarding **vaccines**

Our major sources for facts were from:

World Health Organization

American Academy of Allergy Asthma & Immunology

PublicHealth

RUSH

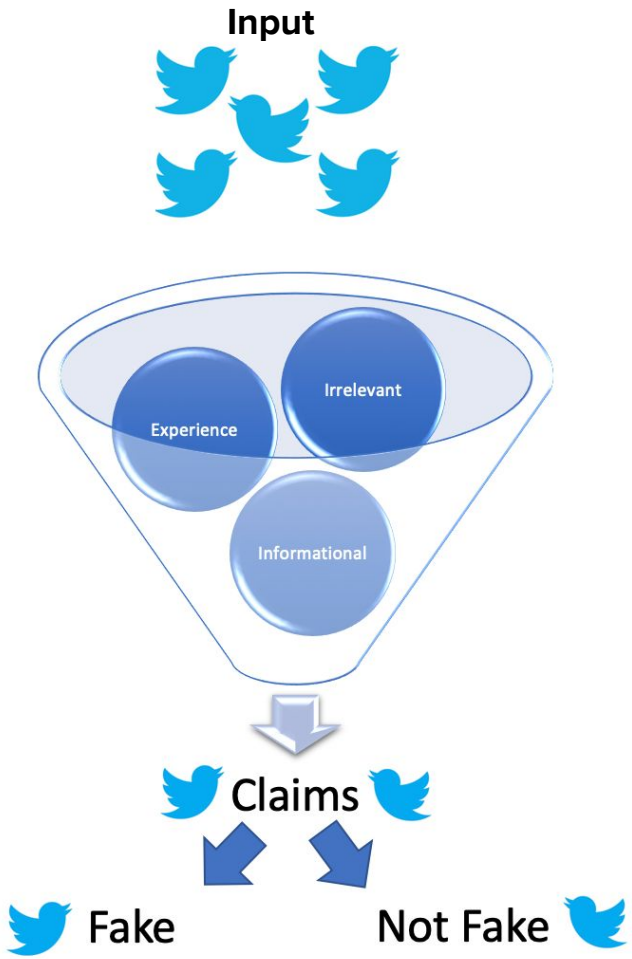
BOSTON PUBLIC HEALTH COMMISSION

love • protect • vaccinate
www.nrvs.info

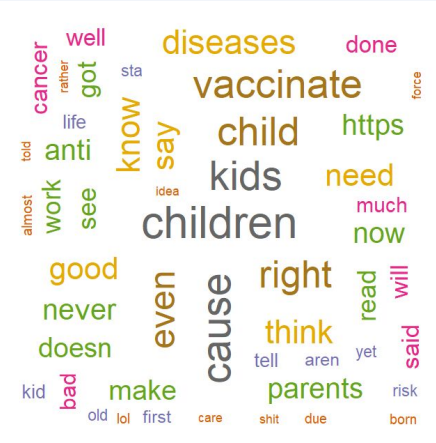
Facts Repository Overview

Myth	Fact
Vaccines cause autism	Symptoms of autism in children develop well before they receive the MMR vaccine
Infant immune systems can’t handle so many vaccines	A baby would theoretically have the ability to respond to around 10,000 vaccines at one time
Natural immunity is better than vaccine-acquired immunity	Vaccine acquired immunity is more efficient than natural immunity
Vaccines contain unsafe toxins	Trace amounts of these chemicals are used in FDA approved vaccines hence safe
Better hygiene and sanitation are actually responsible for decreased infections, not vaccines	The effect of vaccines on curing diseases outweighs sanitation and hygiene
Vaccines aren’t worth the risk	The overall incidence rate of severe allergic reaction to vaccines is usually placed around one case for every one or two million injections.
Vaccines have several damaging and long-term side-effects that are yet unknown. Vaccination can even be fatal.	Vaccines are very safe. Before a vaccine is approved for public use, it is tested thoroughly for many years. Even after vaccines have been approved for the public, vaccines continue to be monitored
Vaccines can infect my child with the disease it’s trying to prevent	The vaccine recipients are experiencing a body’s immune response to the vaccine, not the disease itself

Classification Approach



Manual Classification of Tweets - Buckets



- Classified the tweets into 4 different buckets
- The different buckets are: Claims, Experience, Informational, Irrelevant

Claim - assertions about vaccines and its effectiveness

Experience - testimonials

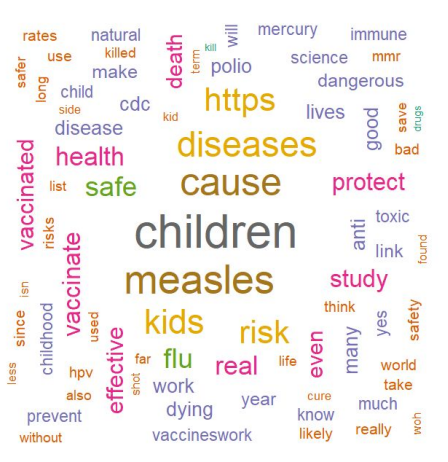
Informational - news regarding vaccination

Irrelevant - none of the above types

Examples:

- very happy to donate during our first #MaxFunDrive! It feels right to thank the good good boys and the rest of the McElroys for hours of laughs, and some tears And as an almost pharmacist I love @Sawbones and vaccines - **experience**
- The #ImmunizeTexas Rally Day is one month away! Will you be there to speak up for #vaccines and fight for science? Visit our website to learn more (link in profile)! #TXLege #VaccinesWork #VaccinesCauseAdults #TIPDay19 - **informational**
- A large new study finds kids who got no childhood vaccines were more likely to be diagnosed with autism than kids who did get recommended vaccinations.- **claim**
- Sul pullman hanno messo ,All my friends are falling in love, dei The Vaccines. CAN U HEAR ME SCREAMING - **irrelevant**

Manual Classification of Tweets - Fake or not



- Initially, we manually classified 1000 Tweets into 4 buckets: “Experience”, “Informational”, “Claims”, and “Irrelevant.”

- We created a model with predictions for the next 500 and used a snowballing technique to ensure those classifications were accurate. We continued that process until 2000 tweets were classified into those 4 buckets.

- All the tweets that were classified as claims (452) were then classified as Fake or Not Fake using our fact repository.

Data Cleaning

Original Tweet- *@JohnHard3 @VirginiaInCal @truthglow @GrainOfSands Glyphosate Found in Childhood Vaccines .... The detection of glyphosate in vaccines ... would indicate the presence of other co-formulants which are also toxic .... Moms Across America sent a letter ... requesting that ....*

- ❏ Changed to lowercase
  - *@johnhard3 @virginiaincal @truthglow @grainofsands glyphosate found in childhood vaccines .... the detection of glyphosate in vaccines ... would indicate the presence of other co-formulants which are also toxic .... moms across america sent a letter ... requesting that ....*
- ❏ Removed usernames
  - *glyphosate found in childhood vaccines .... the detection of glyphosate in vaccines ... would indicate the presence of other co-formulants which are also toxic .... moms across america sent a letter ... requesting that ....*
- ❏ Removed all special characters and numbers
  - *glyphosate found in childhood vaccines the detection of glyphosate in vaccines would indicate the presence of other coformulants which are also toxic moms across america sent a letter requesting that*
- ❏ Removed all stopwords
  - *glyphosate found childhood vaccines detection glyphosate vaccines indicate presence coformulants also toxic moms across america sent letter requesting*

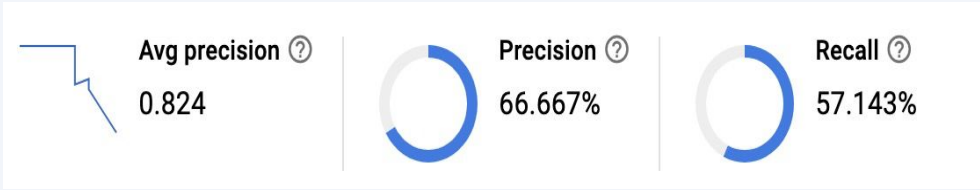
Building Model

- **Buckets**

(Natural Language: Google cloud)

- Built a model on Google cloud AI platform - AutoML Natural Language
- Training data set consists of 1000 manually classified tweets into 4 buckets
- Used the created model API to classify our remaining tweets into buckets

Model score:



Building Model

- **Fake/Not Fake**

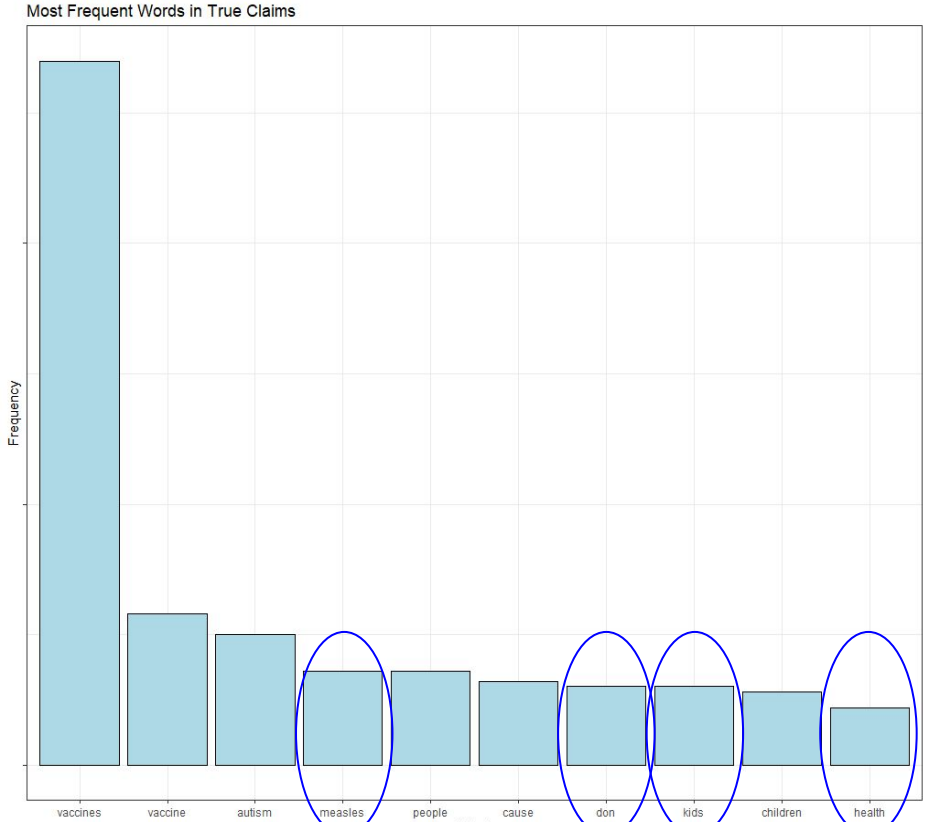
(Naive Bayes Python)

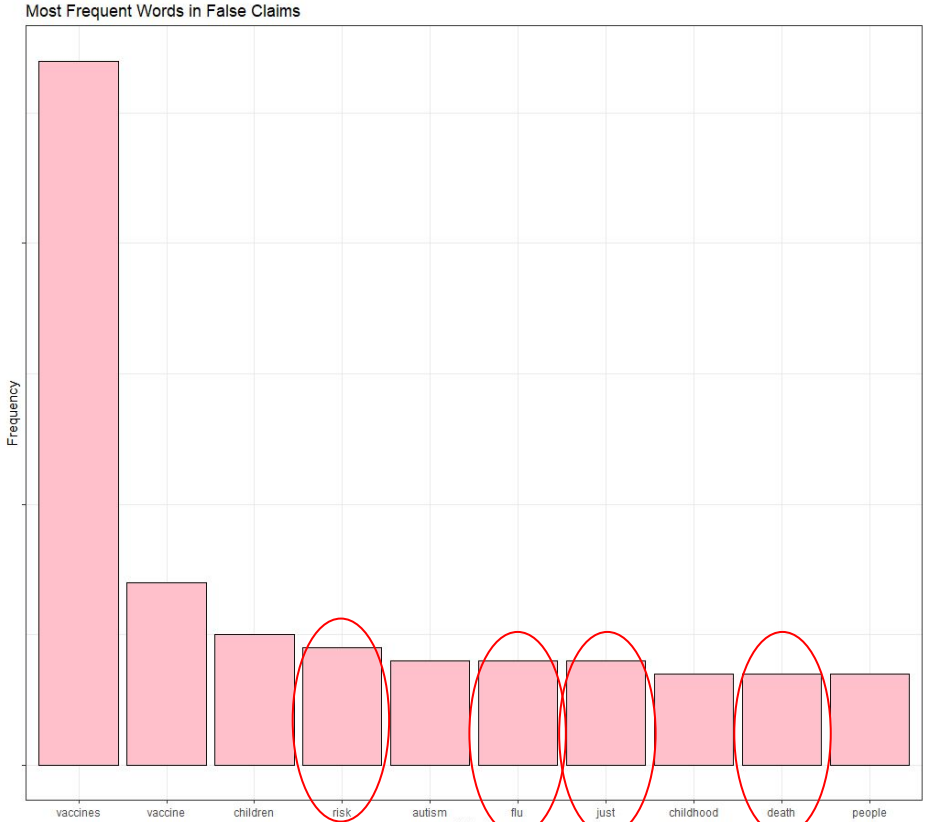
- All tweets labeled “Claims” were used to train the model to predict all unlabeled “Claims” in the Dataset
- The Model used Count Vectorized words as the features and labeled 1 for Fake and 0 for Not Fake.
- Labeled Data was split 80/20 for cross validation
- Naive Bayes Model was used on the vectorized words

Accuracy: 77% Precision: 81%
Recall: 88% F1 Score: 85%

(True = Positive Class)

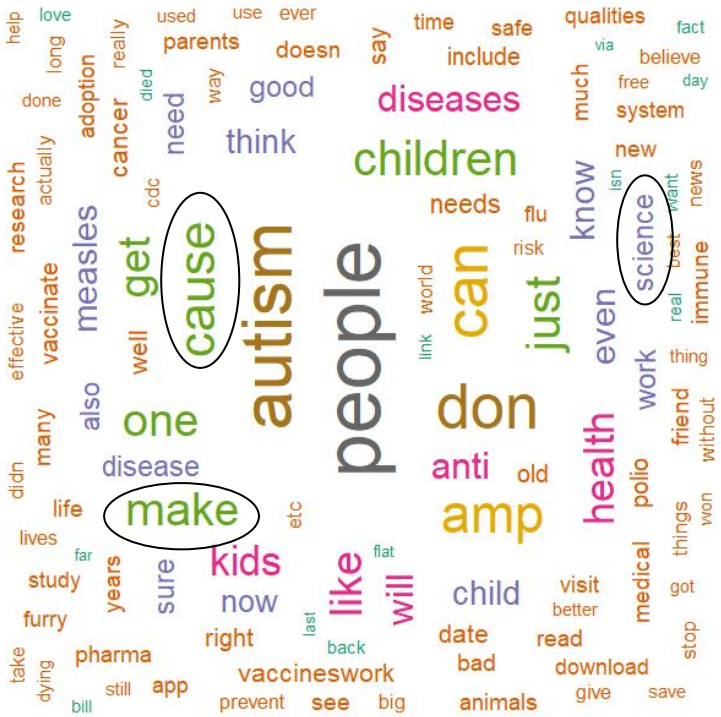
Training Data Word Analysis



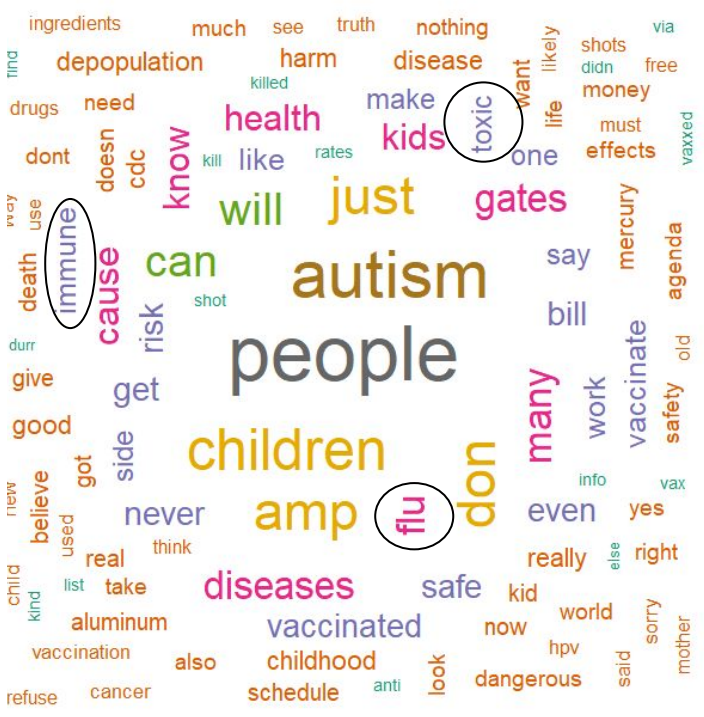


Word Clouds

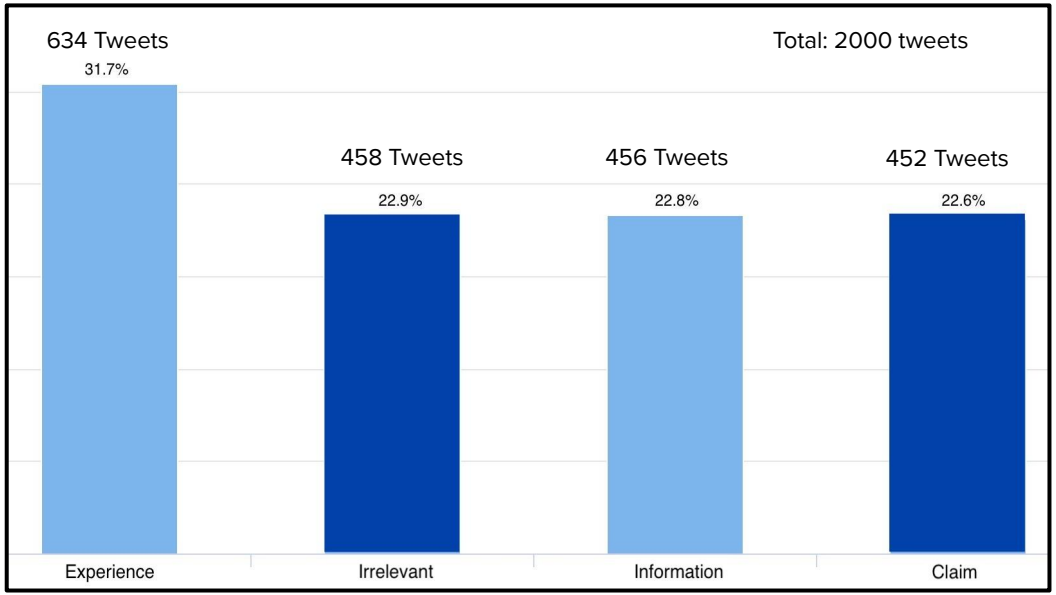
True Claims in the Training Data (1000 Tweets)



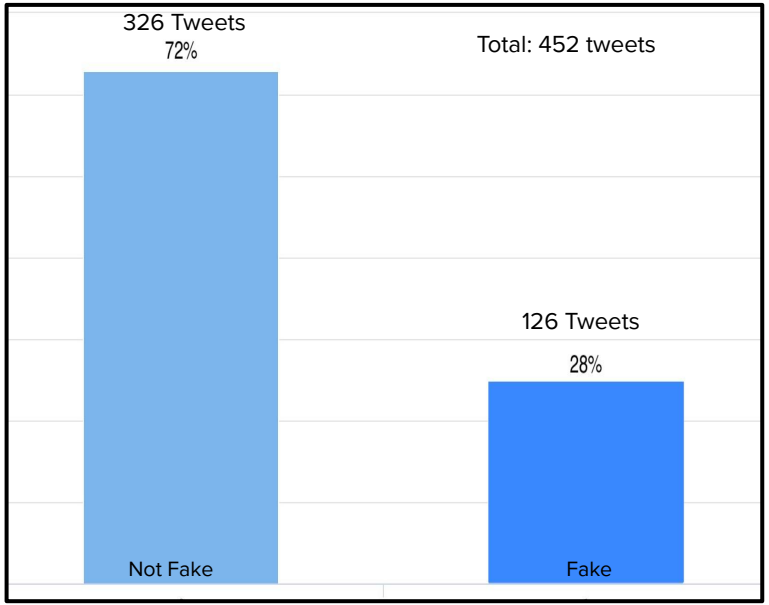
False Claims in the Training Data (1000 Tweets)


people
autism
children
don
amp
just
will
gates
health
kids
toxic
make
immune
cause
can
risk
know
vaccinated
vaccine
diseases
safe
never
real
think
many
work
vaccinate
bill
say
mercury
agenda
safety
old
info
vax
yes
right
really
kid
world
now
hpv
sorry
mother
said
look
dangerous
schedule
anti
childhood
also
vaccination
refuse
cancer
aluminum
take
list
kind
child
believe
used
got
side
good
give
get
death
durr
shot
cdc
doesn
dont
drugs
need
depopulation
harm
disease
nothing
truth
see
much
ingredients
via
shots
didn
free
money
must
effects
vaxxed
life
likely
want
killed
rates
like
one
kill
use
way
mind
new
else
sorry
mother
said
dangerous
look
anti
schedule
cancer
refuse
vaccination
also
childhood
aluminum
vaccinated
now
world
hpv
kid
safe
diseases
really
right
yes
even
vax
info
flu
never
amp
don
children
many
work
vaccinate
bill
say
mercury
agenda
safety
old
autism
people
get
give
good
side
got
believe
used
child
kind
list
take
real
think
cause
risk
shot
immune
can
will
just
gates
kids
toxic
make
health

Analysis - Distribution

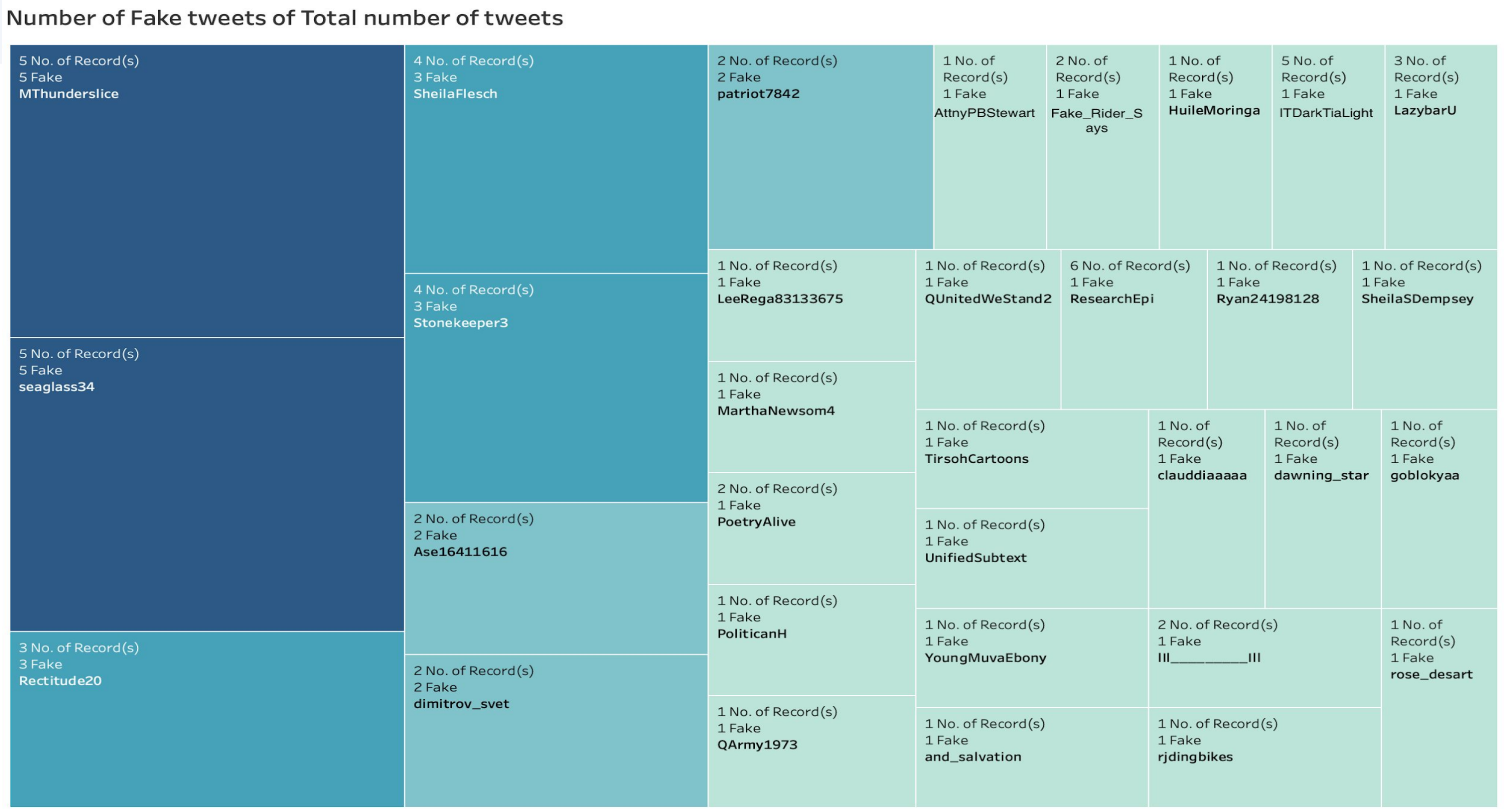


Distribution of Buckets



Distribution of Fake in Claims

Analysis - Fake tweets/Total no. of tweets

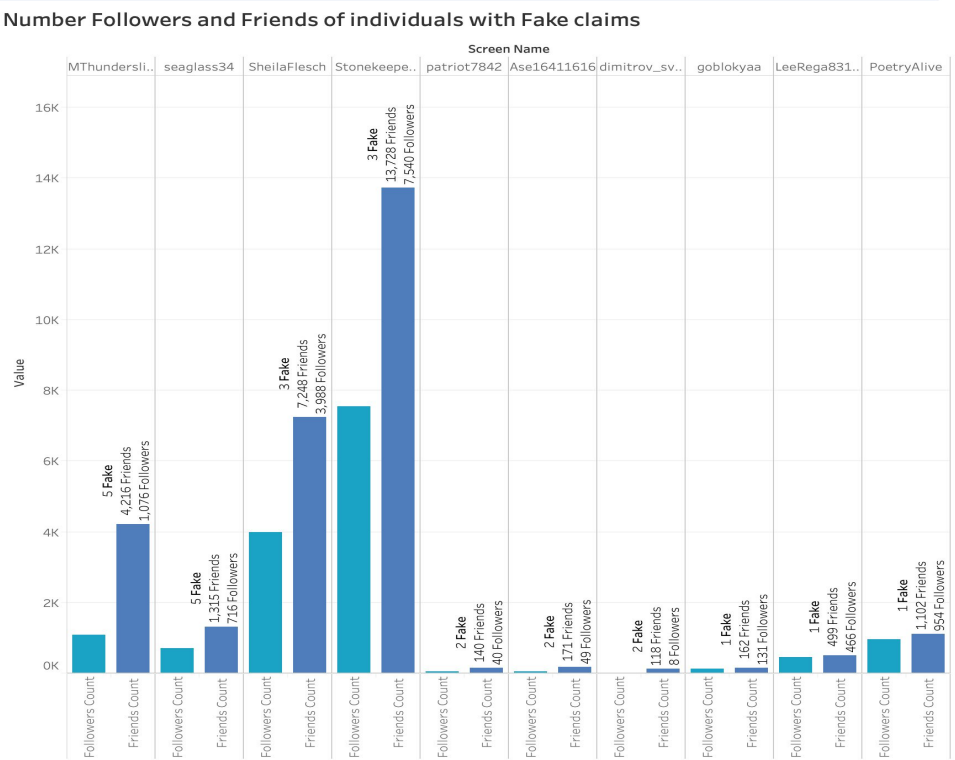


Analysis of Fake claims

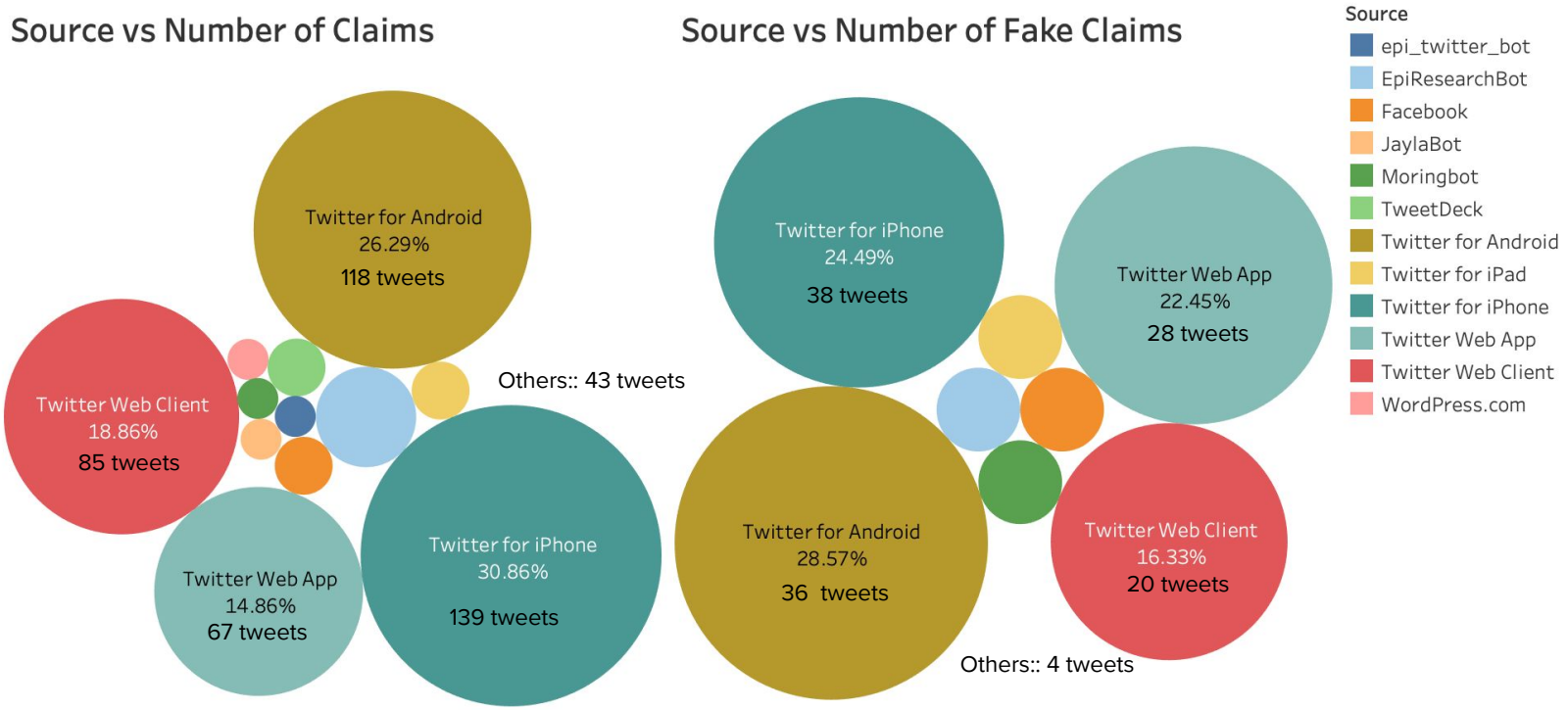
How many fake claims were Retweeted?

Is Retweet	
False	61.22%
True	38.78%

False : 77 Tweets
True : 49 Tweets



Analysis - Source of tweets



Result - snippet

Fake Claims

Example 1: not only are many vaccines filled with toxins but gardasil is dangerous unnecessary amp ineffective it has not decreased cervical cancer but sure has harmed tens of thousands of the people you are supposed to protect

Example 2:a large new study finds kids who got no childhood vaccines were more likely to be diagnosed with autism than kids who did get recommended vaccinations

Non Fake Claims

Example 1:the hpv vaccine is most effective during adolescence when the body produces more antibodies against the virus thats why recommends the vaccine at ages 11 or 12

Example 2:modern vaccines are highly effective and safe they do not cause autism or contain toxic chemicals

Thank You