

NATURAL LANGUAGE PROCESSING

Fake News Detection

Jasmine Vasandani



Tips to help prevent the spread of rumors & fake news*

1. Understand when a message is forwarded
2. Check photos and media carefully
3. Look out for messages that look different
4. Check your biases
5. Fake news often goes viral
6. Verify with other sources
7. Help stop the spread

Support The Guardian

Available for everyone, funded by readers

Contribute →

Subscribe →

Search jobs

Sign in

Search

The Guardian

US edition ▾

News

Opinion

Sport

Culture

Lifestyle

More ▾

US

World

Environment

Soccer

US Politics

Business

Tech

Science

WhatsApp

This article is more than 1 month old

WhatsApp 'deleting 2m accounts a month' to stop fake news

App launches paper on 'stopping abuse' in India, home to more than 200m of its users

Michael Safi in Delhi

@safimichael

Wed 6 Feb 2019 11:00 EST

f t e

< 276



Editorially independent, open to everyone

We chose a different approach – will you support it?

Support The Guardian →

Source

Wired

How WhatsApp Fuels Fake News and Violence in India

SIGN IN | SUBSCRIBE

TIMOTHY McLAUGHLIN

BACKCHANNEL

12.12.18 07:00 AM

HOW WHATSAPP FUELS FAKE NEWS AND VIOLENCE IN INDIA

SHARE

f SHARE 876

t TWEET

COMMENT

EMAIL

BY THE TIME police arrived in the hamlet of Rainpada on July 1, 2018, the village council office was the scene of a massacre.

THE BRIGHT BLUE shutters on the windows were splintered and the door was kicked in. Inside, files were strewn across the floor and the light green walls were splashed with blood.

GET WIRED
AD-FREE BROWSING
+ UNLIMITED ACCESS

SUBSCRIBE

MOST POPULAR

Source

How might Natural Language Processing better inform us about fake news?




CASE STUDY:

Distinguish between **fake news** and **absurd news** by building a classification model trained on Subreddit posts from r/TheOnion and r/nottheonion.

r/TheOnion

COMMUNITY DETAILS

 **r/TheOnion**

95.0k
Subscribers

79
Online


Articles from The Onion. This is not /r/nottheonion. Only links to the Onion (and its satirical sister sites) are allowed here. [ClickHole](http://www.clickhole.com/) and [StarWipe](http://www.starwipe.com/) are the Onion's sister sites, and links to them are welcome here.

SUBSCRIBE

CREATE POST

r/nottheonion

COMMUNITY DETAILS

 **r/nottheonion**

15.0m
Readers

9.9k
Online

For true stories that are so mind-blowingly ridiculous that you could have sworn they were from The Onion.

SUBSCRIBE

CREATE POST

Methodology

1

DATA ACQUISITION

Scrape 15k posts from r/TheOnion & r/nottheonion, total of 30k posts. Used pushshift.io API wrapper to acquire data.

3

NATURAL LANGUAGE PROCESSING

Prepare data for modeling. Remove punctuation & #s. Drop duplicates. All lowercase.

2

EXPLORATORY DATA ANALYSIS

Create data visualizations to observe trends about the data. What are distinguishing characteristics about each subreddit?

4

CLASSIFICATION MODELING

Find the best combination of vectorizers and models to achieve the best accuracy when designating a post as either from r/TheOnion or r/nottheonion.

Data Acquisition

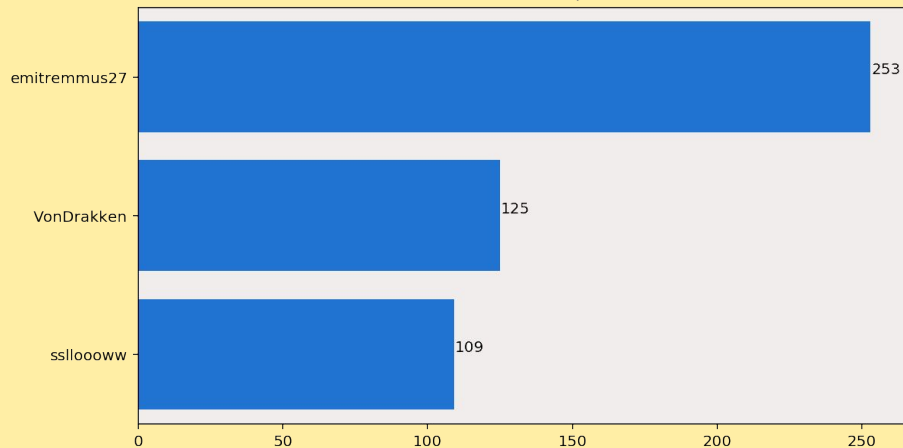
- **29,867 posts**, 15k from r/nottheonion, rest r/TheOnion
 - After duplicates, 26,275 posts
 - Features: title (predictor), subreddit (target), comments, author, # sub, score, domain, date
- **Posts Date Range:**
 - r/TheOnion: Oct 4, 2013 - Apr 7, 2019
 - r/nottheonion: Jan 25, 2019 - Apr 7, 2019
- **Data Cleaning:**
 - numbers
 - punctuation
 - case

“

The first biggest mistake in data science is looking [at] every business challenge as [a] predictive approach. Remember 70% of low hanging problems can be solved by just doing an **EDA**.

Sundar Ramamurthy

Most Active Authors: r/nottheonion



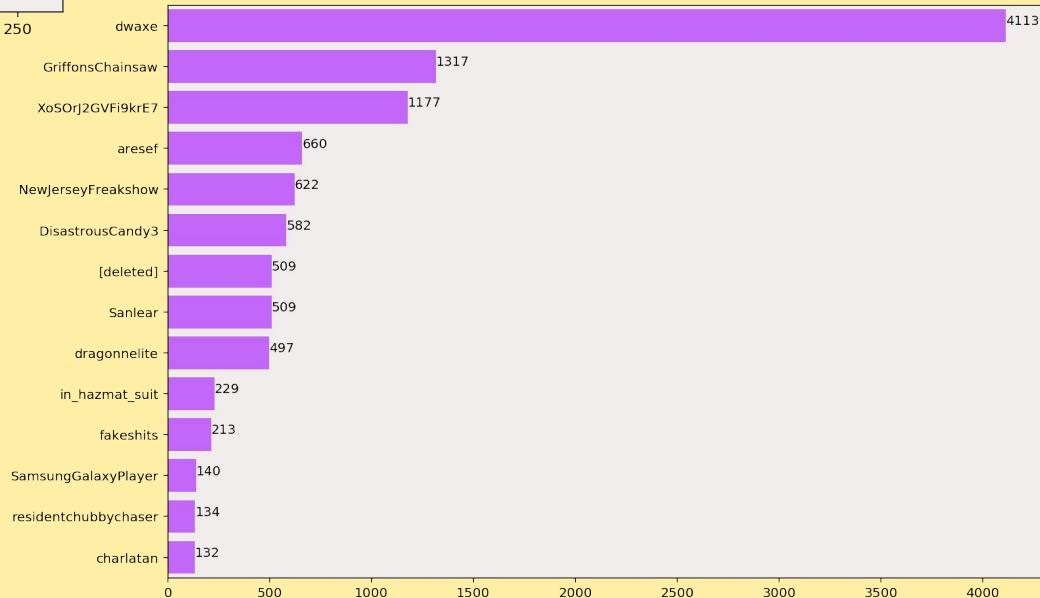
ACTIVE AUTHORS | r/nottheonion

15m subscribers, only 3 authors shared over 100 posts.

ACTIVE AUTHORS | r/TheOnion:

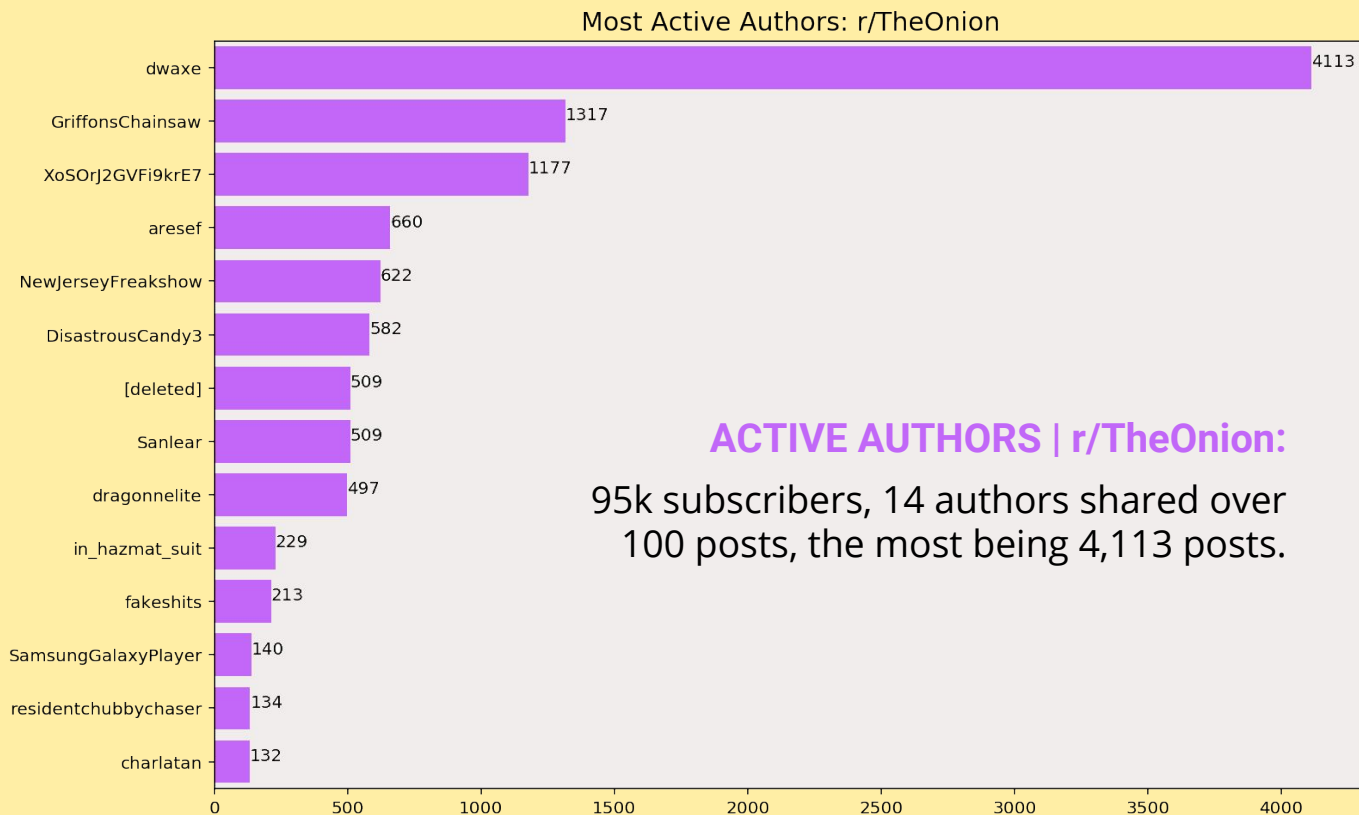
95k subscribers, 14 authors shared over 100 posts, the most being 4,113 posts.

Most Active Authors: r/TheOnion



FAKE NEWS PREVENTION VIOLATIONS:

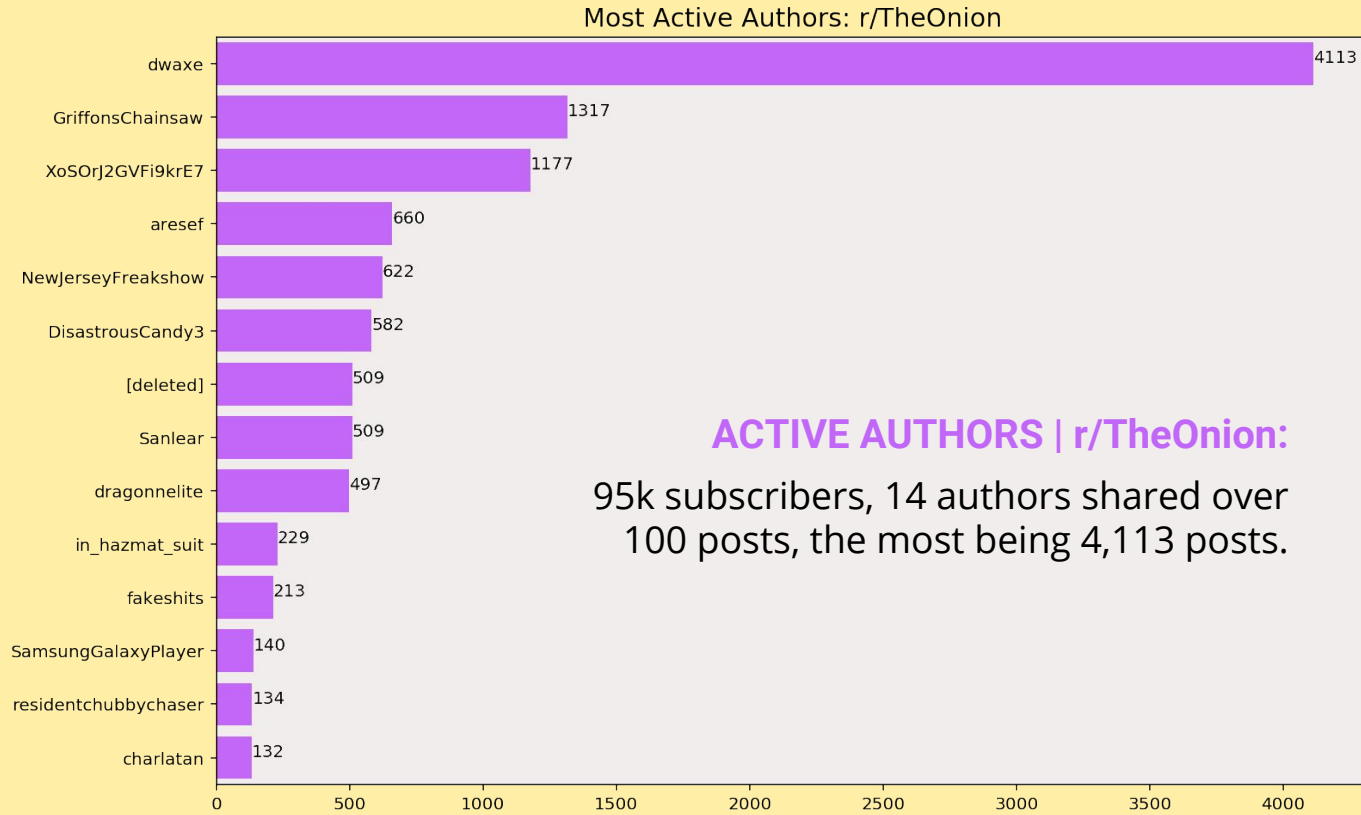
- Understand when a message is forwarded
- Fake news often goes viral
- Help stop the spread



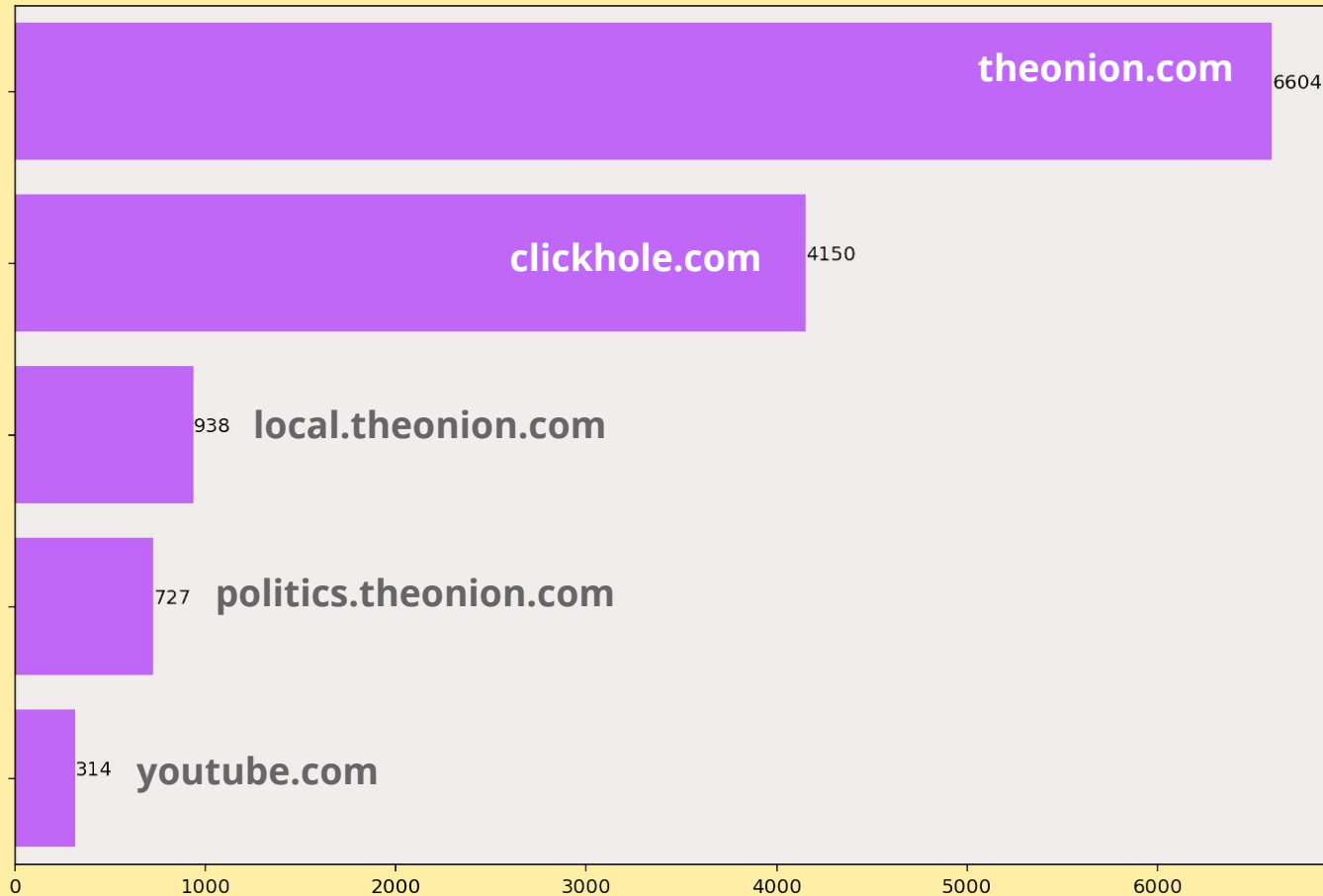
FAKE NEWS PREVENTION VIOLATIONS:

- Understand when a message is forwarded
- Fake news often goes viral
- Help stop the spread

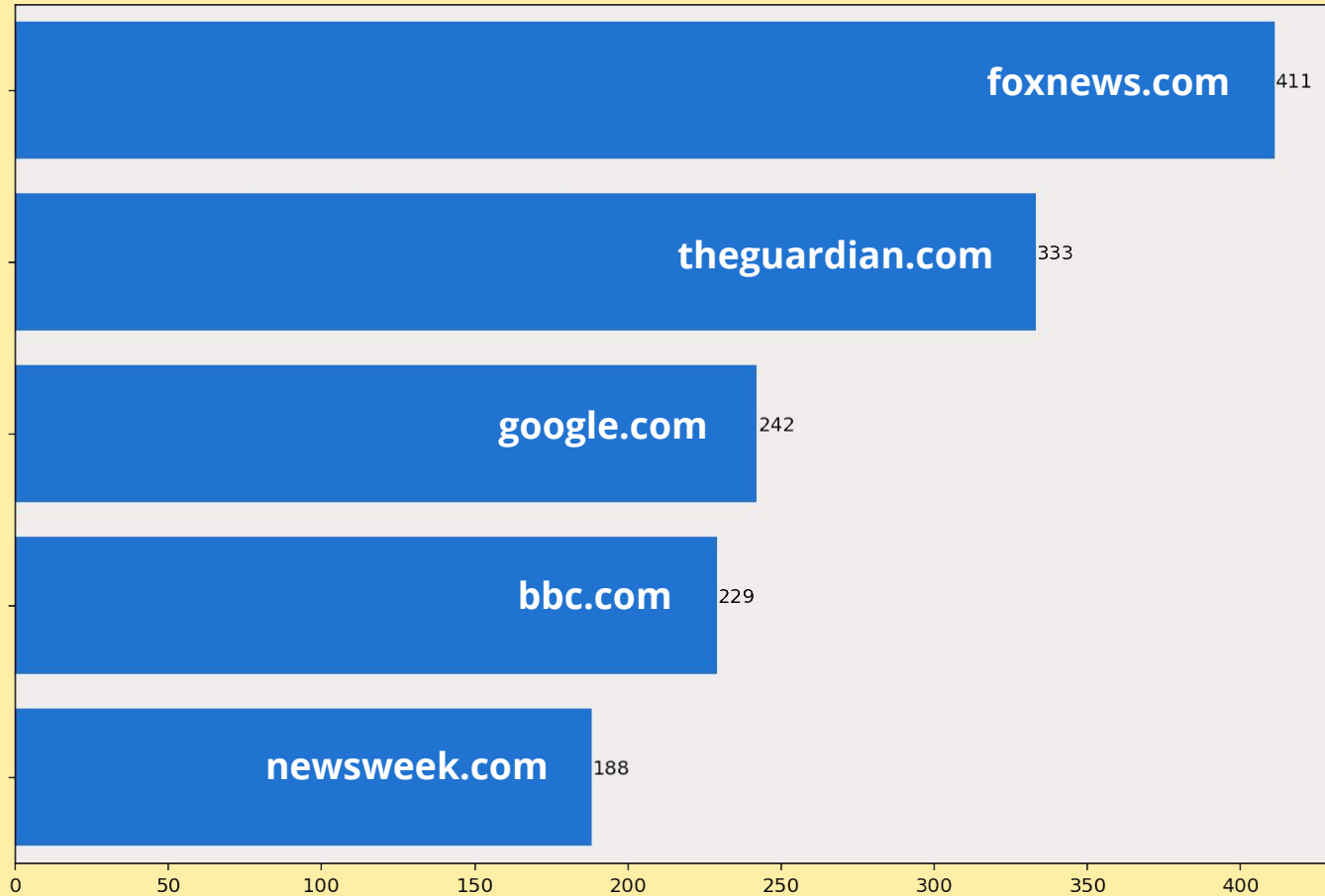
r/TheOnion is a good case study to analyze fake news trends.



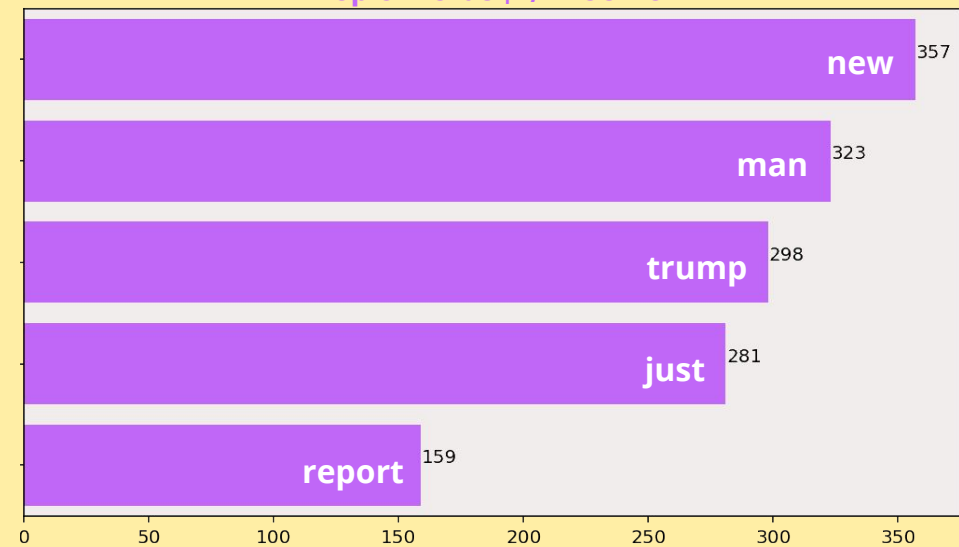
TOP 5 MOST REFERENCED DOMAINS | r/TheOnion



TOP 5 MOST REFERENCED DOMAINS | r/nottheonion



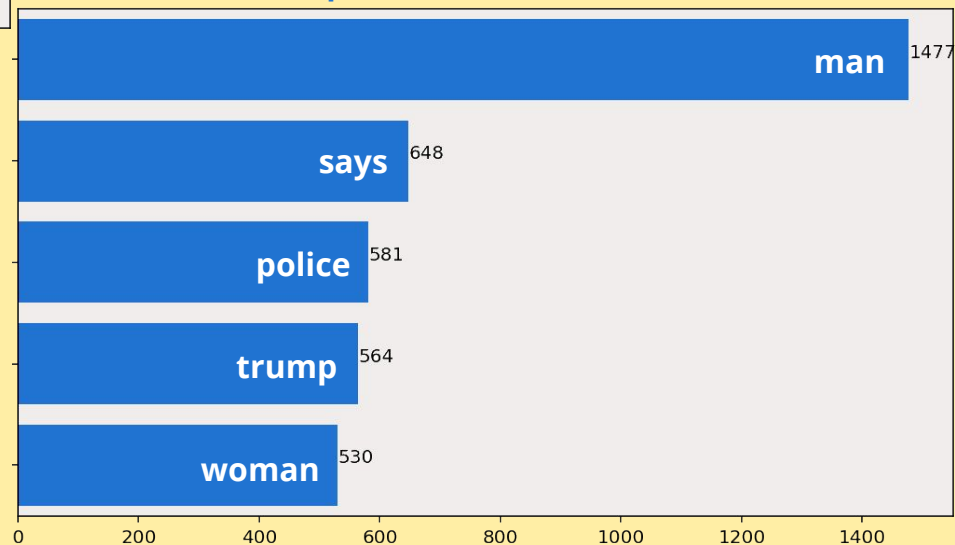
Top 5 Words | r/TheOnion



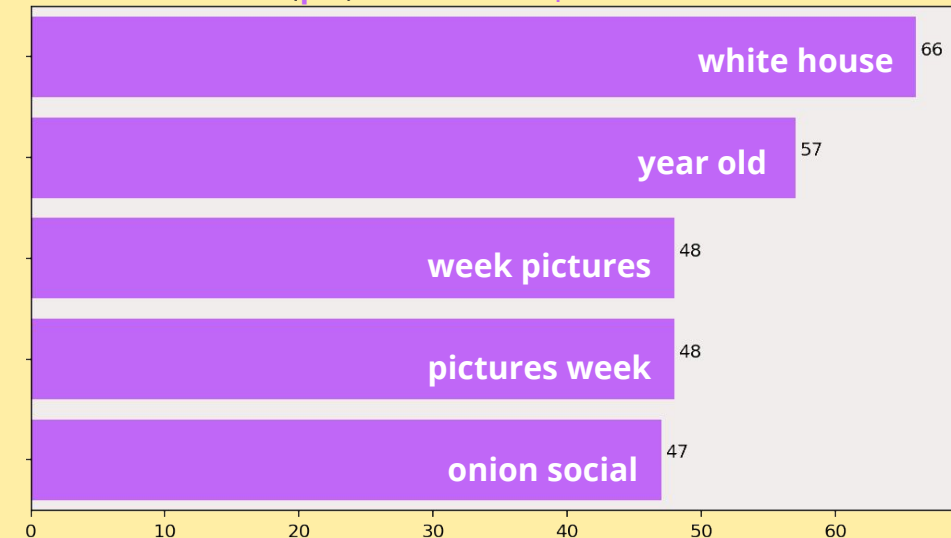
MAN

TRUMP

Top 5 Words | r/nottheonion



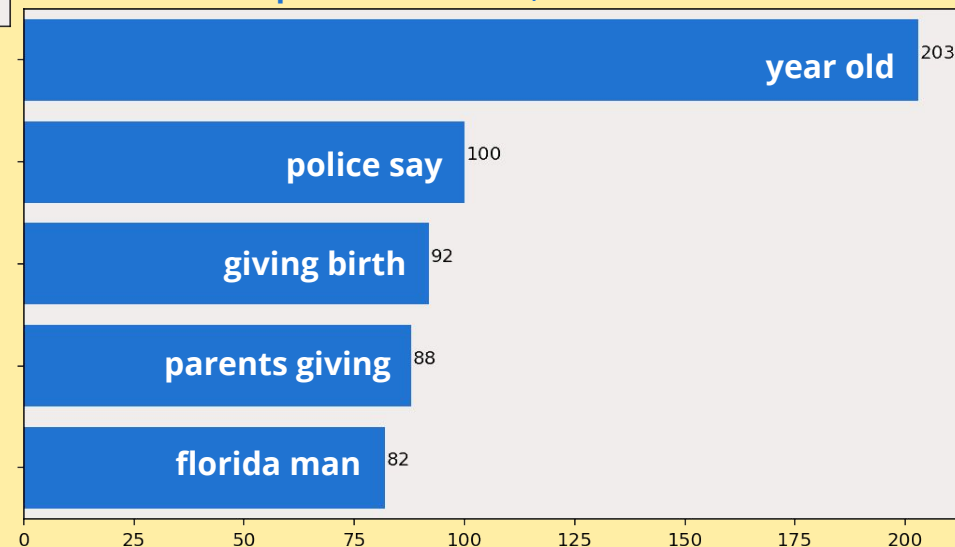
Top 5 Paired Words | r/TheOnion



YEAR

OLD

Top 5 Paired Words | r/nottheonion



MODELING

Pipeline & GridSearchCV

CountVectorizer()

stop_words: None, 'english', custom
ngram_range: (1,1), (2,2), (1,3)

LogisticRegression()

C: 0.01, 1

CountVectorizer()

stop_words: None, 'english', custom
ngram_range: (1,1), (2,2), (1,3)

MultinomialNB()

C: 0.01, 0.36, 1

TfidfVectorizer()

max_df: .75, .98, 1.0
min_df: 2, 3, 5
ngram_range: (1,1), (1,2), (1,3)

LogisticRegression()

C: 1

TfidfVectorizer()

max_df: .75, .98, 1.0
min_df: 4, 5
ngram_range: (1,1), (1,2), (1,3)

MultinomialNB()

C: 0.1, 1, 100

Best Accuracy Score: 90%

CountVectorizer()

stop_words: None, 'english', custom
ngram_range: (1,1), (2,2), (1,3)

LogisticRegression()

C: 0.01, 1

BEST TEST SCORE: 0.87

Train score: 0.96

CountVectorizer()

stop_words: None
ngram_range: (1,3)

MultinomialNB()

C: 0.36

BEST TEST SCORE: 0.90

Train score: 0.997

TfidfVectorizer()

max_df: .75, .98, 1.0
min_df: 2, 3, 5
ngram_range: (1,1), (1,2), (1,3)

LogisticRegression()

C: 1

BEST TEST SCORE: 0.86

Train score: 0.90

TfidfVectorizer()

max_df: .75, .98, 1.0
min_df: 4, 5
ngram_range: (1,1), (1,2), (1,3)

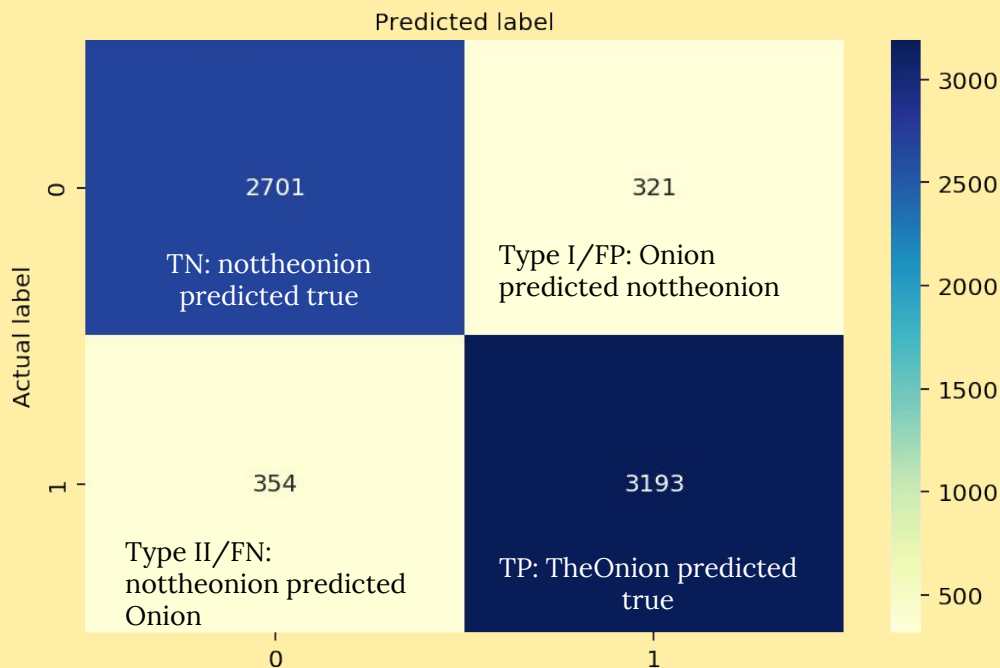
MultinomialNB()

C: 0.1, 1, 100

BEST TEST SCORE: 0.86

Train score: 0.92

Model 3: Confusion Matrix



Accuracy

For all predictions, 90% predicted correctly.

Positive Predictive Rate

Among all Onion predictions, 91% predicted correctly.

True Positive Rate

Among all Onion titles, 90% predicted correctly.

True Negative Rate

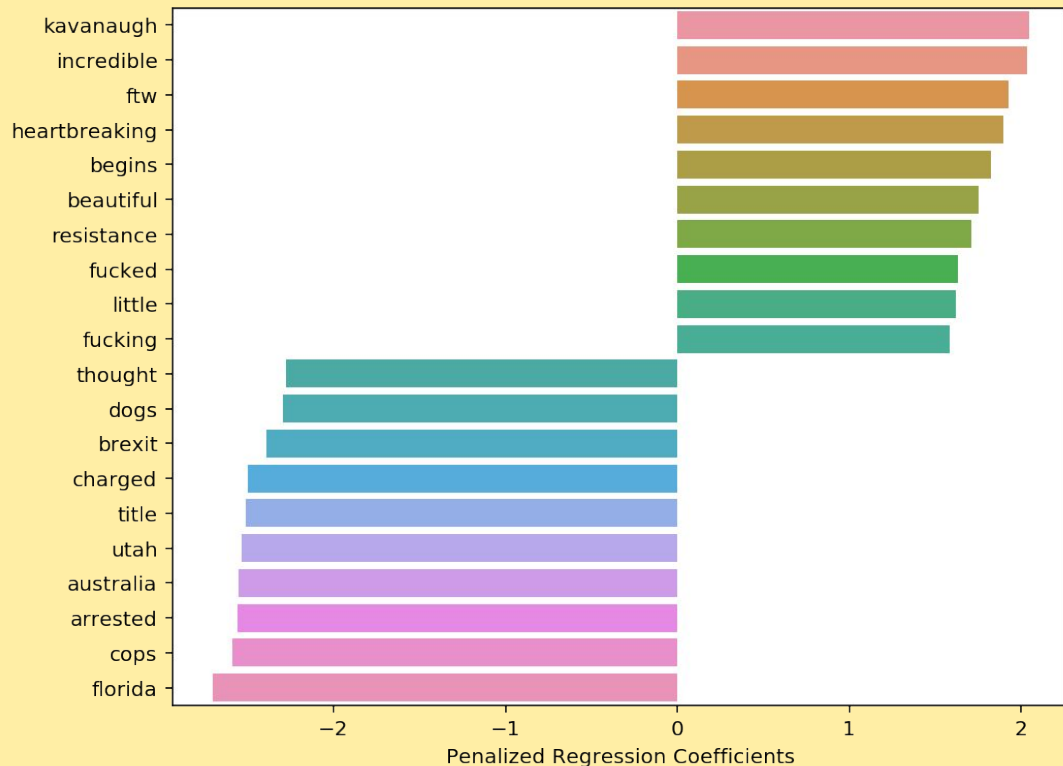
Among all nottheonion, 89% predicted correctly.

Misclassification Rate

For all predictions, 11% predicted incorrectly.

Logistic Regression Coefficients

MODEL 3



r/TheOnion

The word that contributes the most positively to being from r/TheOnion is 'kavanaugh', followed by the 'incredible' and 'ftw'.

- As occurrences of "kavanaugh" increases by 1 in a title, that title is 7.7 times as likely to be classified as r/TheOnion.

r/nottheonion

The word that contributes the most positively to being from r/nottheonion is 'Florida', followed by 'cops' and 'arrested'.

- As occurrences of "florida" increases by 1 in a title, that title is 14.9 times as likely to be classified as r/nottheonion.

Reflections

- NLP of text is one way to analyze fake news, but a **major gap** exists:
 - *image & video analysis*

r/IAmA

- Ask me anything!



[Source.](#)