Detecting Fake News

Aditya Koduri

Real vs Fake

If there's anyone in this day and age who most definitely doesn't mince words, and will say what needs to be said, it's **Vice President Joe Biden**. **While speaking in Philadelphia**, Pennsylvania on a **campaign stop** for Hillary Clinton, Biden tore into some of the remarks from Monday's debate. Especially when Donald Trump seemed to think it's perfectly fine to screw over the American people, or benefit from their sorrow. Biden said: "This is a guy who said — and wants to be president — that it was good business for him to see the housing market fail. What in the hell is he talking about?" Adding: **"Every president I have served with, including Republicans, has had a moral center** about what it was to be an American, about what we're supposed to do, about what basic fundamental rights are." Then really getting to the core of it, he said: "He does not have the basic fundamental sensibilities and values that almost every American politician left, right and center I know have. They disagree on how to make things better for you, but they don't take pleasure from 'You're fired.'— NBC Nightly News

"The nose on the Hillary Clinton double at the apartment is clearly different than all pictures of Hillary in the past. The different nose is a dead giveaway, in the identification of nose types, the real **Hillary Clinton** has what is known as the Nordic nose, which is quite common and straight. The **clone** has a distinctively different nose that slightly curves downward at the tip and has a slight elevated area along the bridge, this being more of a Roman type. Aside from the different nose types, **the clone Hillary rather appears to be a younger version of the original with much less aging apparent**. Hillary's stand in also has no mole above her right lip, Hillary Clinton does. Since the **real Hillary Clinton obviously has such bad health issues that practically prevents her from even campaigning**, we must assume that her clone stand in will do most of the appearances she is scheduled to do before the election, aside from those Bill Clinton will do himself. **What we need to look for immediately is her appearance. If she appears to be much thinner, younger, or without a cough and hoarseness, we should know, this is not the real Hillary Clinton.** If she appears to be right on the mark, without any incidents of brain freeze or forgetfulness, this is the clone, not Hillary Clinton.

Real vs Fake

If there's anyone in this day and age who most definitely doesn't mince words, and will say what needs to be said, it's **Vice President Joe Biden**. While campaigning in Philadelphia, Pennsylvania on a **campaign stop** for Hillary Clinton, Biden tore into some of the remarks from Monday's debate. Especially when Donald Trump seemed to think it's perfectly fine to screw over the American people tomorrow. Biden said: "This is a guy who said — and wants to be president — that it was good business for him to see the housing market fail. What in the hell is he talking about?" Adding: "**Every single thing I've ever done, with, including Republicans, has had a moral center** about what it was to be an American, about what we're supposed to do, about what basic fundamental rights are." Then really getting to the heart of it, he said: "I don't ever hate the basic fundamental sensibilities and values that almost every American politician left, right and center I know have. They disagree on how to make things better for you, but they don't take pleasure from 'You're fired.'— NBC Nightly News

Real news

- Can be opinionated

- Reader can disagree

- **Can get a few facts wrong

"The nose on the Hillary Clinton double at the apartment is clearly different than all pictures of Hillary in the past. The different nose is a dead giveaway, in different types, the real **Hillary Clinton** has what is known as the Nordic nose, which is quite common and straight. The **clone** has a distinctively different nose that slightly curves downward at the tip and has a slight elevation at the bridge more of a Roman type. Aside from the different nose types, **the clone Hillary rather appears to be a younger version of the original with much less aging apparent**. Hillary's stand in at the apartment looks younger than the Hillary Clinton. Since the **real Hillary Clinton obviously has such bad health issues that practically prevents her from even campaigning**, we must assume that her clone stand in will do most of the appearances and scheduled work before the election, aside from those Bill Clinton will do himself. **What we need to look for immediately is her appearance. If she appears to be much thinner, younger, or without a cough and hoarseness, we should know, this is not the real Hillary Clinton.** If she appears to be right on the mark, without any incidents of brain freeze or forgetfulness, this is the clone, not Hillary Clinton.
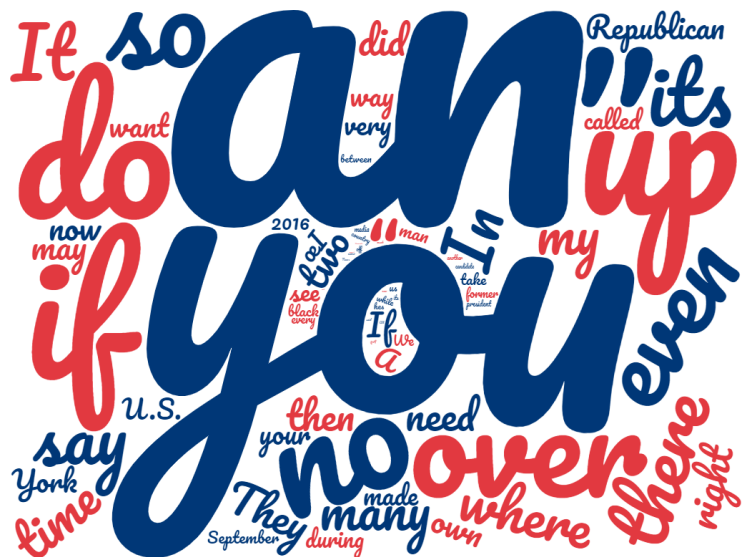
Fake news

- Strongly opinionated

- Reader disagreement is shunned

- Based on a false premise

Data Overview - Buzzfeed

Field	Description
ID	Unique ID for new article
source	New source for Buzzfeed
date	Date of the article
title	Article title
text	Corpus of text
fake	0 or 1
date_time	Timestamp

Date range: 19th to 28th Sep 2016 (68% in 10 days and 27% missing date)

33 sources

Source	Source %	Fake news %
http://politi.co	18%	0%
http://rightwingnews.com	13%	71%
http://cnn.it	13%	0%
http://eaglerising.com	10%	61%
http://freedomdaily.com	9%	94%

Data Processing

Pre-processing is required!!

Use NLTK libraries

- Convert to lowercase
- Remove stop words –the, is, which,…
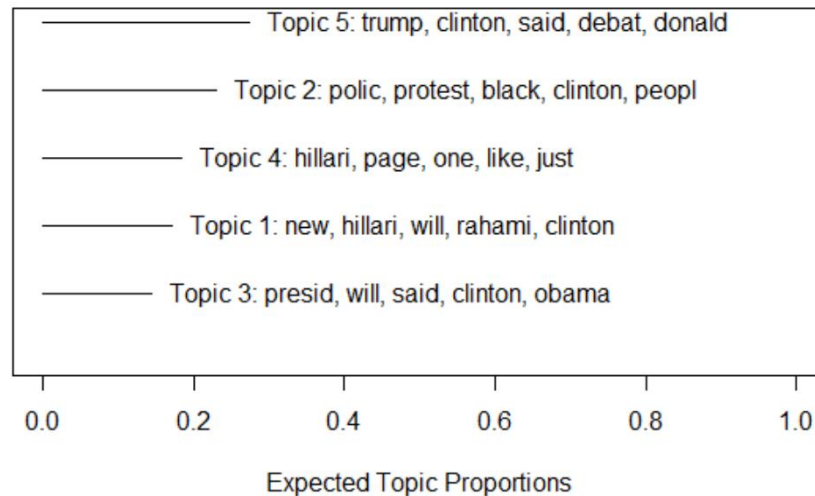- Remove special chars - punctuation, numbers
- **Stemming & Lemmatization**

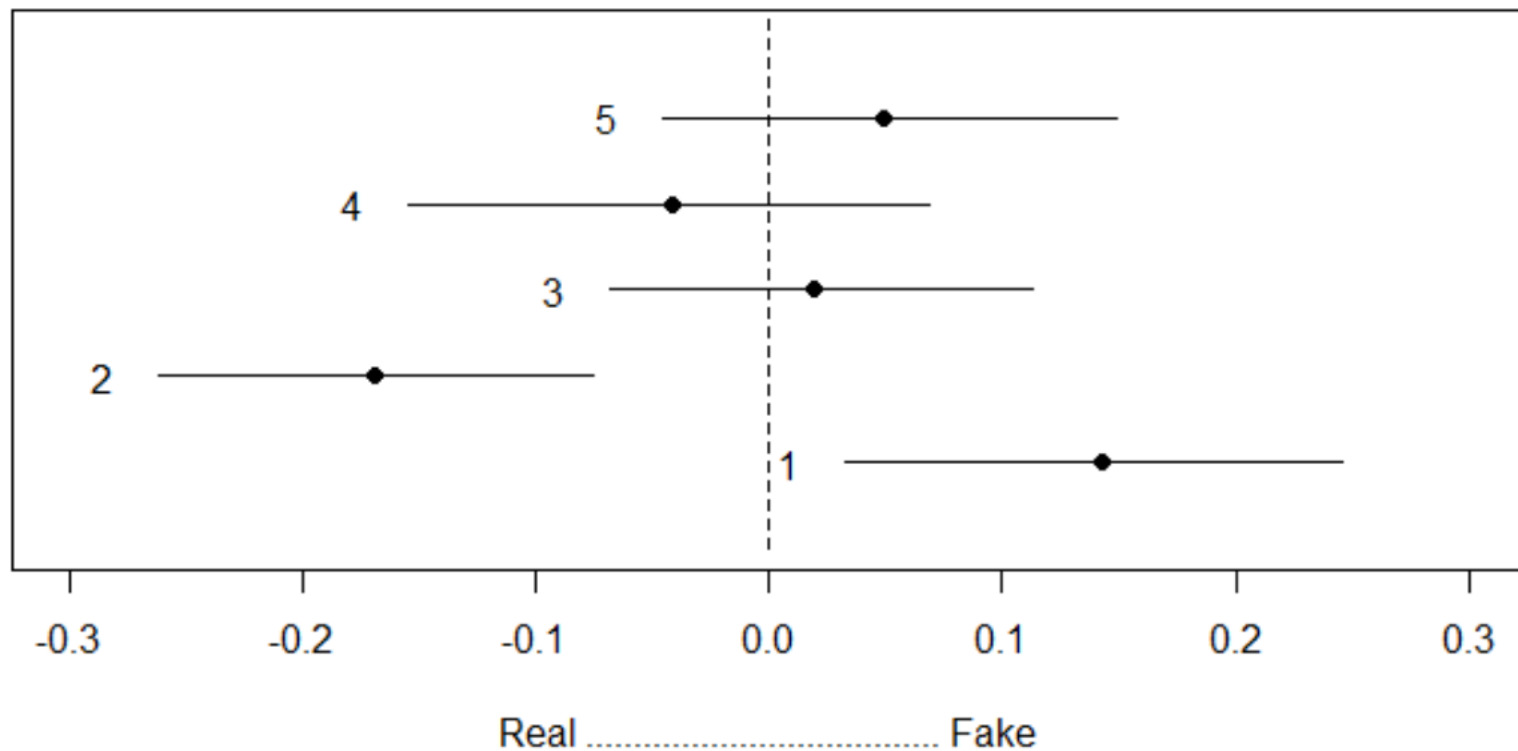| Word	Stemming	Lemmatization
Walking	Walk	Walk
Cars	Car	Car
Caring	Car	Care
Stripes	Strip	Stripe

Topic Modelling

- Linear Discriminant Analysis (LDA) with Expectation Maximization (EM)
- 2 Simulations with 5 & 10 topics

Topic 1
christi, rahami, feder, port, arpaio, testifi, counti, birth, document, citizenship

Topic 2
charlott, ranbaxi, explos, protest, chai, polic, islam, terror, riot, muslim

Topic 3
crisi, charit, nuclear, biden, global, war, kennedi, migrant, reagan, percent

Topic 4
lesson, mint, hawkin, page, moscow, advic, adult, russian, e-mail, fbi

Topic 5
cruz, conway, poll, moder, debat, trump, tax, fact-check, donald, holt

Top Topics

Topic 5: trump, clinton, said, debat, donald

Topic 2: polic, protest, black, clinton, peopl

Topic 4: hillari, page, one, like, just

Topic 1: new, hillari, will, rahami, clinton

Topic 3: presid, will, said, clinton, obama

0.0   0.2   0.4   0.6   0.8   1.0

Expected Topic Proportions

Effect of Real vs. Fake

Next Steps

Gradient boost classifier : 70%

SVC : 86.11%

Random Forest : 90%

XGBoost : 80.5%

```
model = RandomForestClassifier(n_estimators=500, max_depth=15,random_state=9)
model.fit(X, label)
y_predictions = model.predict(X_test)
print("test-set score: {:.3f}".format(f1_score(np.array(test.fake), y_predictions)))
print("test-set score: {:.3f}".format(accuracy_score(np.array(test.fake), y_predictions)))

test-set score: 0.914
test-set score: 0.903
```