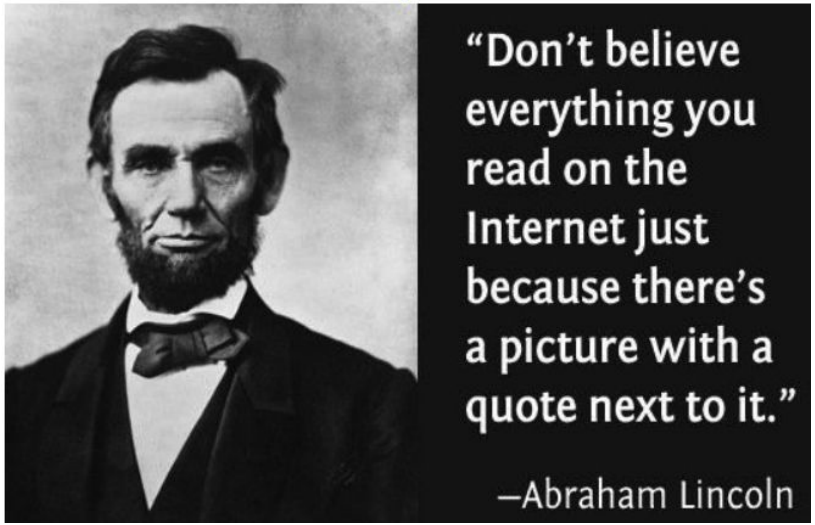
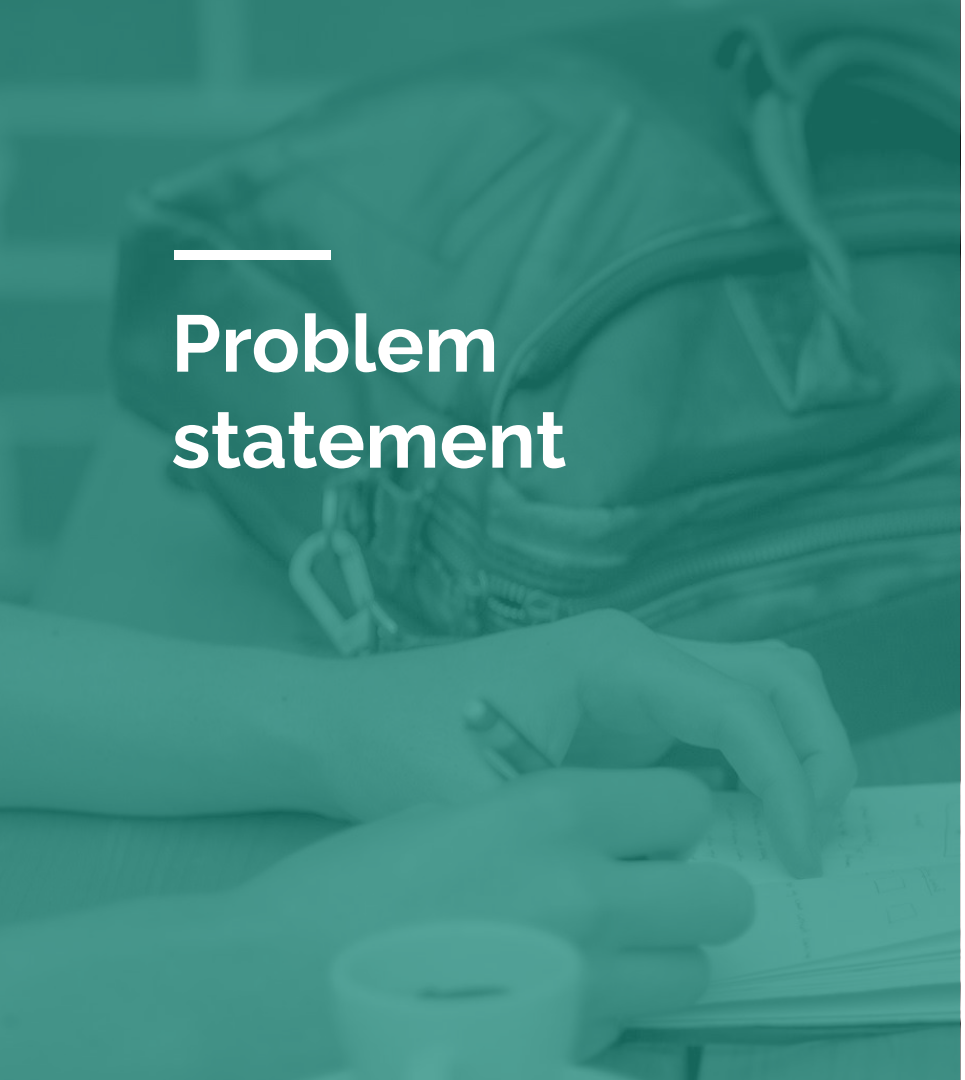

CPAC
NATIONAL HARBOR, MD
10:32 AM ET
AMERICAN
CONSERVATIVE
UNION
CPAC
PRES TRUMP: FAKE NEWS IS THE ENEMY OF THE PEOPLE
FOX NEWS ALERT
FOX


“Don’t believe everything you read on the Internet just because there’s a picture with a quote next to it.”
—Abraham Lincoln

Fake News Detector

By Georgie Zaguirre



Problem statement

Why is Fake News is a problem?

The spread of disinformation and hoaxes to deliberately deceive and attract website hits are now accelerated via social media.

Exploitation of social media to try to influence elections.

Gartner predicts that By 2022, majority of individuals in mature economies will consume more false information than true information.

Impact to Businesses

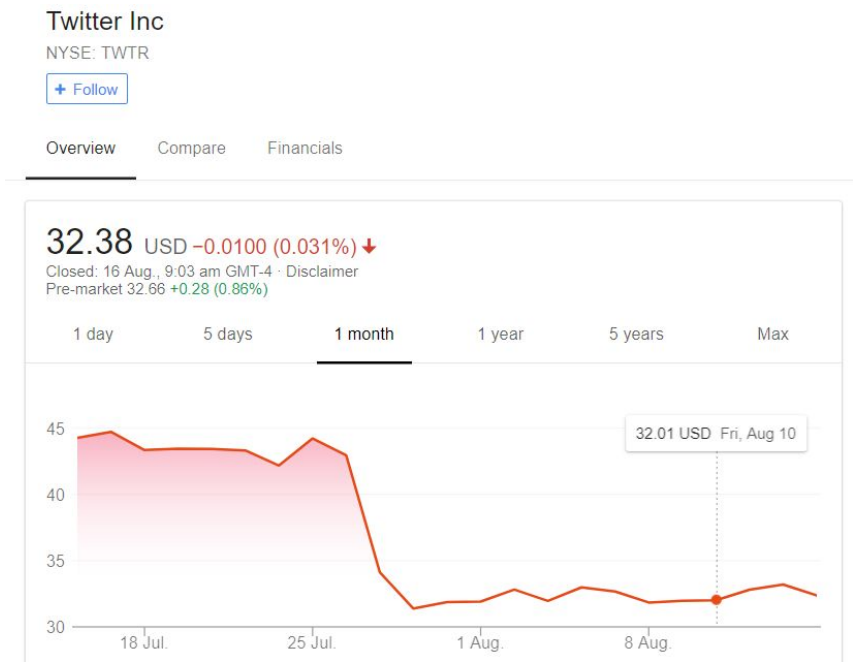
Facebook has recently removed 583 million fake accounts, as a result of new privacy laws and scandals .

$160 billion wiped from its market value

Impact to Businesses

So did Twitter…

$6 billion wiped from its market value



What is being done today?

Society, governments and organisations are only now starting to recognise and act on the problem

- France has passed law to ban Fake News
- UK has an ongoing parliamentary inquiry to examine Fake News

Who will benefit from a Fake News Detector?

BBC NEWS

Fairfax Media

Project Goals

- Predict Fake News articles using Natural Language Processing (NLP)
- Perform Sentiment Analysis to predict whether a news article is positive, negative or neutral using unsupervised learning

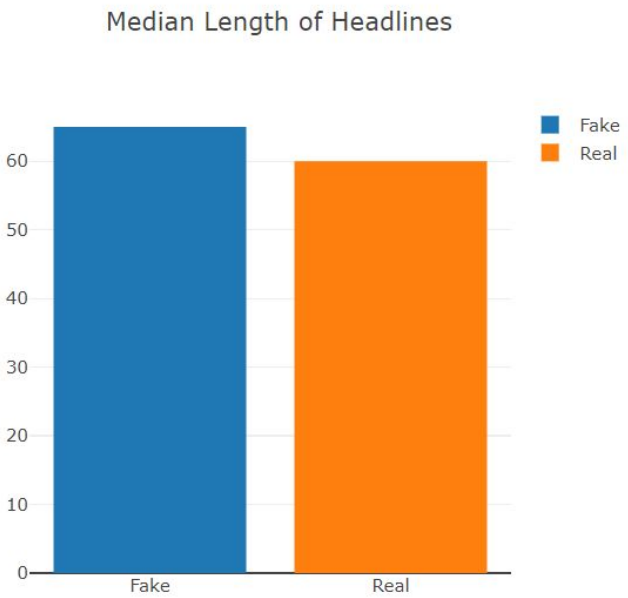
Prediction Steps

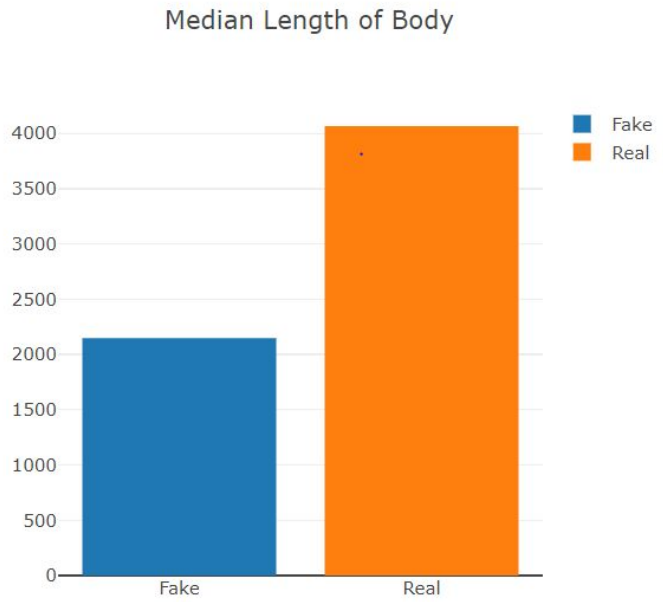
- ➔ Source Data
- ➔ Cleaning, pre- processing
- ➔ Exploratory Data Analysis
- ➔ Select learning algorithm (prediction model)
- ➔ Model Evaluation

The Dataset

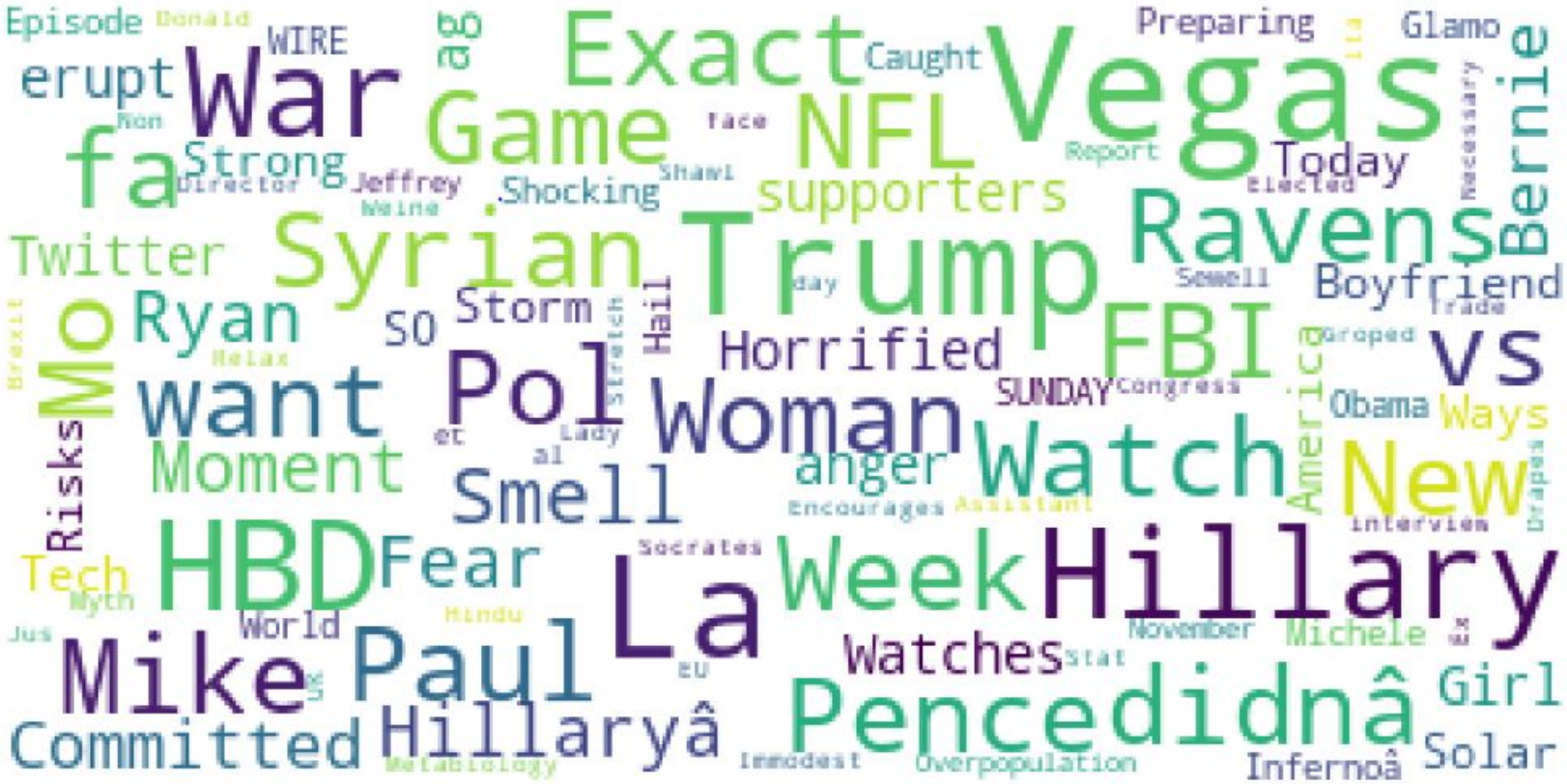
- 10154 articles
- Sourced from kaggle, datacamp
- CNN, BBC, Reuters, Fox News
- Beforeitsnews, Change.org, Moveon.org

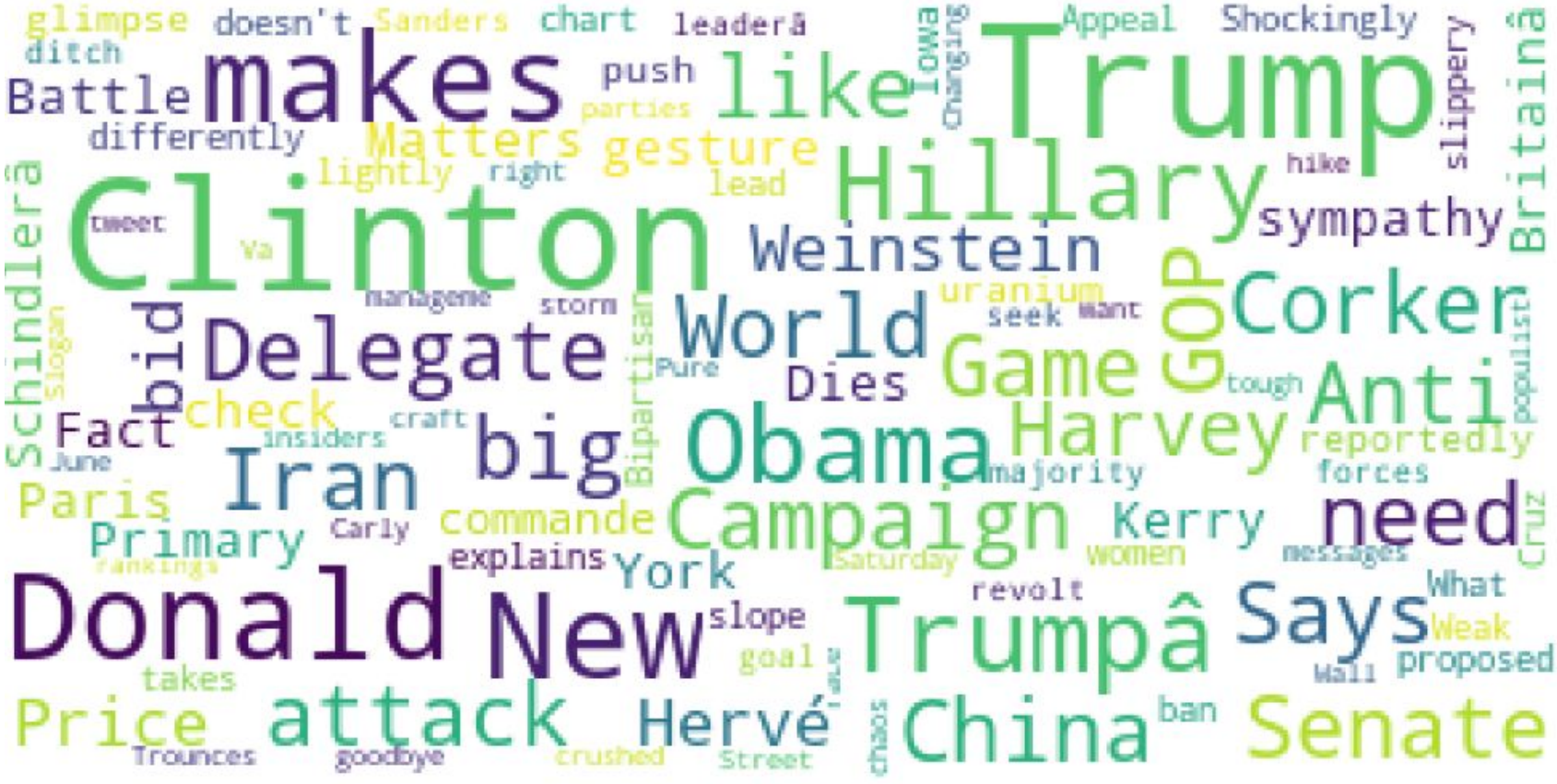
Length


Median Length of Headlines
Fake
Real
60
50
40
30
20
10
0
Fake
Real


Median Length of Body
Fake
Real
4000
3500
3000
2500
2000
1500
1000
500
0
Fake
Real

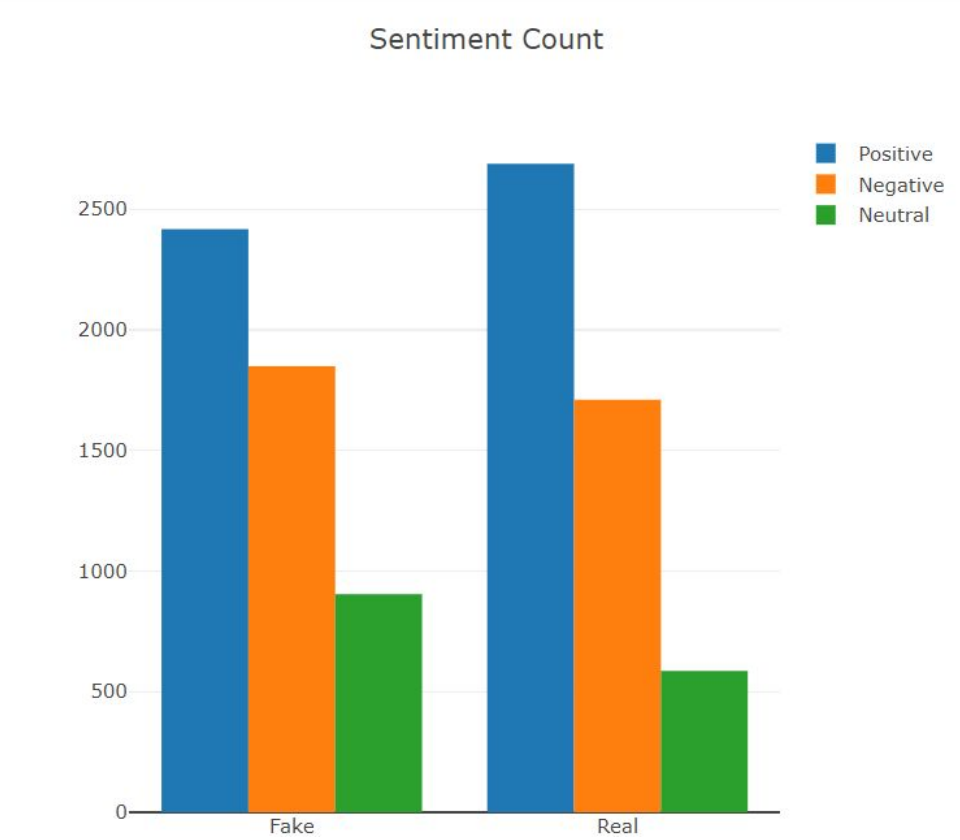
Fake News headlines frequent words

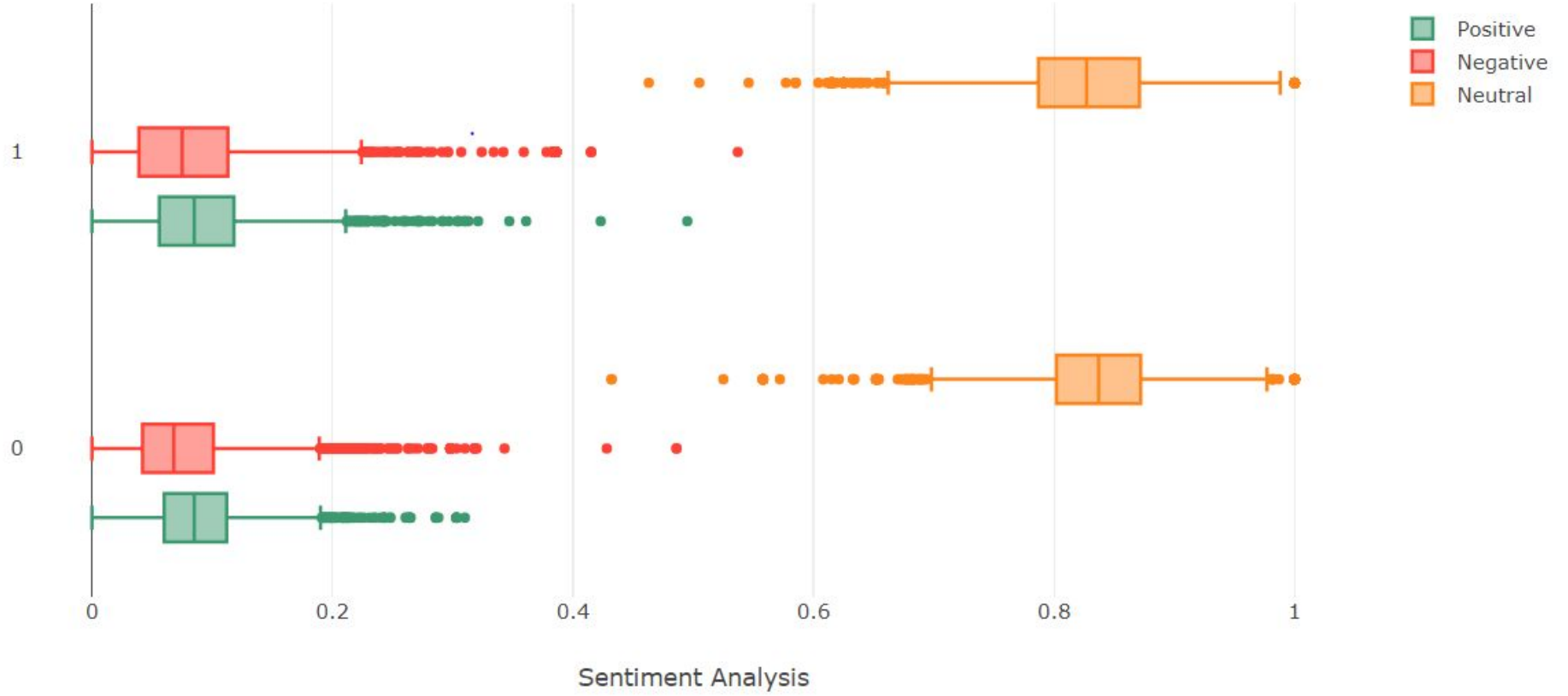

Episode
Donald
WIRE
ag
Exact
Caught
Preparing
Glamo
erupt
War
Vegas
Bernie
Game
NFL
Report
Today
fa
Strong
Director
Jeffrey
Shocking
supporters
Ravens
Twitter
Syrian
Trump
Boyfriend
Ryan
Storm
SO
Hail
Horrified
FBI
vs
Mo
want
Pol
Woman
SUNDAY
Congress
Obama
Ways
Risks
Moment
Smell
anger
Watch
New
America
HBD
Fear
Week
Hillary
Tech
Myth
Socrates
La
Mike
Paul
World
Watches
November
Michele
Girl
Committed
Hillaryâ
Pencedidnâ
Solar
Infernoâ
Innodest
Overpopulation

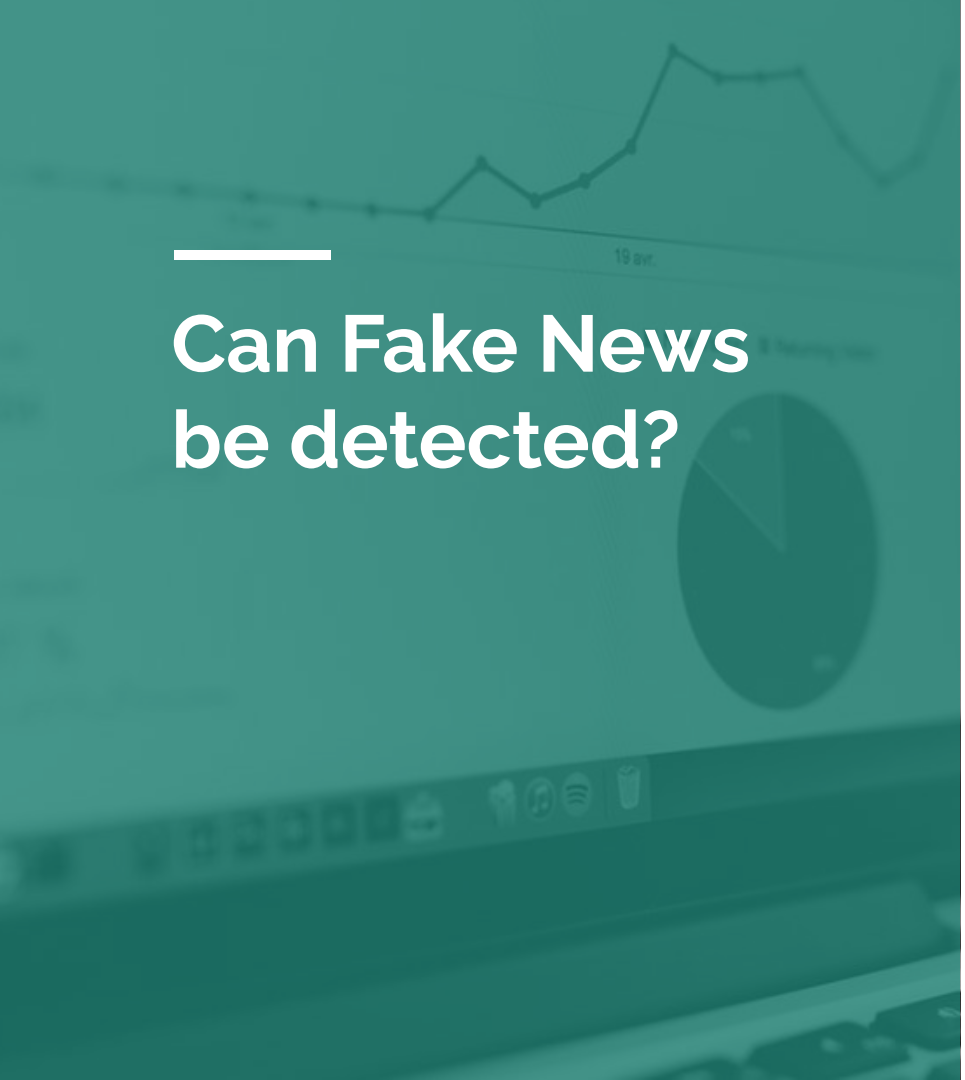

glimpse
doesn't
Sanders
chart
leaderâ
Appeal
Shockingly
ditch
Battle
makes
push
like
Iowa
changing
Trump
Britaina
parties
differently
Matters
gesture
slippery
hike
lightly
right
lead
Schindlerâ
Clinton
Hillary
tweet
Va
sympathy
Weinstein
uranium
GOP
Corker
storm
seek
want
bid
Delegate
World
Game
Anti
Pure
Dies
tough
check
Fact
craft
Obama
Harvey
reportedly
insiders
June
Iran
big
Bipartisan
majority
forces
Paris
Campaign
Kerry
need
Primary
Carly
commande
Cruz
explains
York
Saturday
women
messages
rankings
revolt
What
Says
Donald
New
slope
Trumpâ
Weak
goal
takes
proposed
Wall
Price
attack
Hervé
China
ban
Senate
Trounces
goodbye
crushed
Street
chaos

Sentiment Analysis

Can the model correctly predict if a news article is either Positive, Negative or Neutral?


Sentiment Count
Positive
Negative
Neutral
2500
2000
1500
1000
500
0
Fake
Real


Positive
Negative
Neutral
1
0
0
0.2
0.4
0.6
0.8
1
Sentiment Analysis



Can Fake News be detected?

The simple answer is Yes!

The model can predict Fake News with 93% accuracy

Summary of Findings

- Fake News headlines are longer in length than Real News article
- Real News body of text are nearly twice as long as Fake News articles
- Headlines or frequent words are similar between Fake News and Real News
- There are more Positive counts of news articles than Negative and Neutral
- The overall sentiment with both Real News and Fake News are generally Neutral
- The model can predict Fake News with **93% accuracy**

Assumptions, limitations and risks

This model assumes that Real News are articles published by traditional media organisations such as BBC, CNN and Fox News. Whereas Fake News are published by less known or non-traditional news organisations such as Moveon.org, Beforeitnews.com and Change.org

There is a risk that the model is biased towards Fake News and certain topics. Many of the news articles in the dataset were written during the 2016 US Election campaign - a time when there was a large spread of propaganda and misleading information for political purposes.

Since this is a binary classification, the model does not go as far as predicting bias, satire or conspiracy. It assumes that these classes are Fake News

Next Steps

What can be done next?

Web app

Expand the model to predict multiple classes

Expand data to include wider topics such as science, sports environment etc and not just politics

Scrape data from social media as Facebook and Twitter

Questions?