



# Detecting Fake News

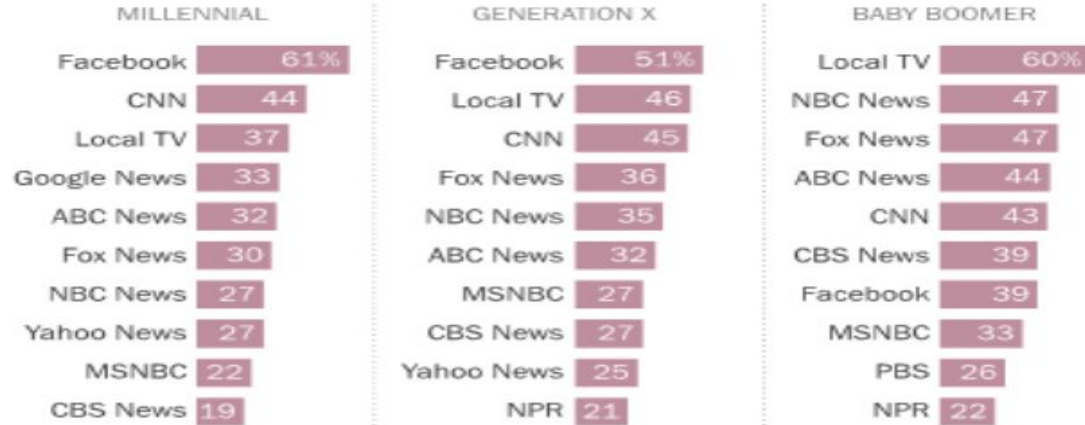
Prepared by Ji Hoon Chung

# Overview & Business Problem

Client : Social Network Platforms (Facebook, Twitter, LinkedIn, etc.)

## Among Millennials, Facebook Far Exceeds Any Other Source for Political News

*% who got news about politics and government in the previous week from...*

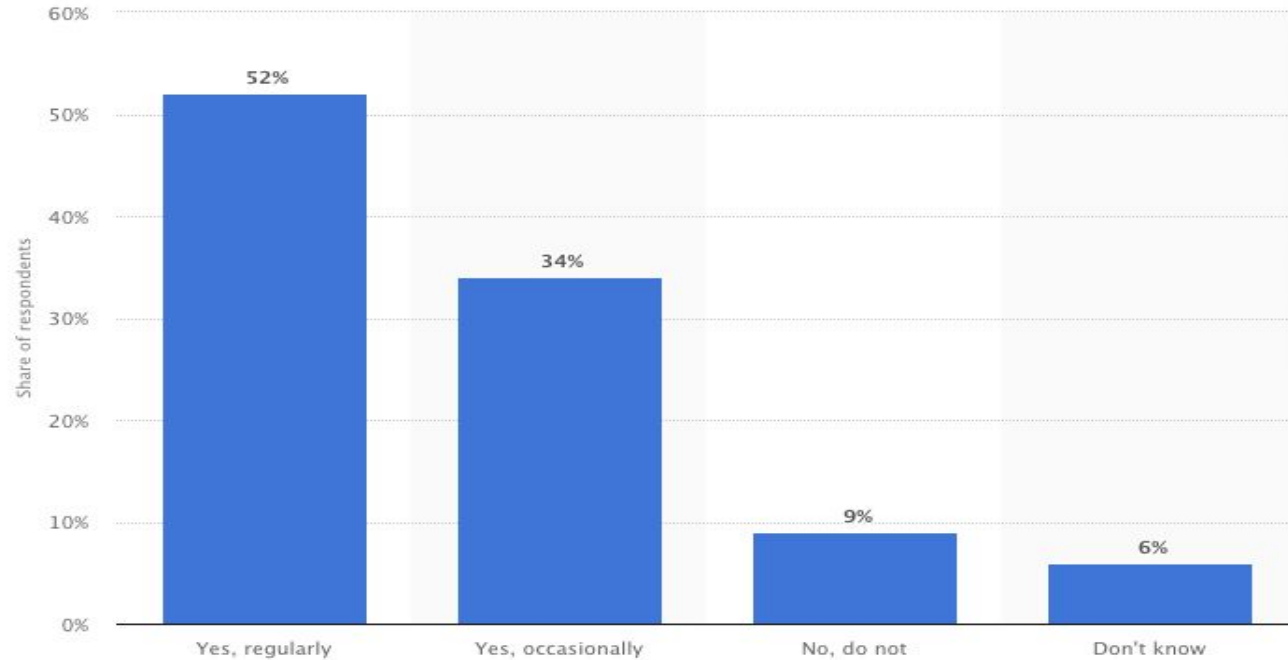


American Trends Panel (wave 1). Survey conducted March 19-April 29, 2014. Q22, Q24A.  
Based on online adults.

PEW RESEARCH CENTER

# Overview & Business Problem

Client : Social Network Platforms (Facebook, Twitter, LinkedIn, etc.)



Source: Statista 2021



# Data Exploration

What is FAKE NEWS?

“Fake news is a type of yellow journalism or propaganda that consists of deliberate misinformation or hoaxes spread via traditional print and broadcast news media or online social media.”

FAKE news Examples:

*Here, we examine the claim that boiled garlic will “cure” the virus.*

*Joe Biden Calls Trump Supporters 'Dregs of Society’*



# Data Exploration

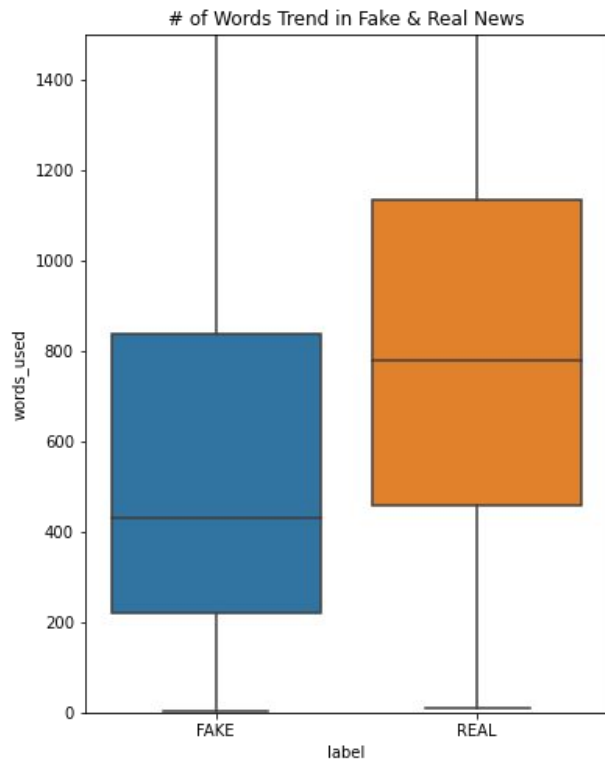
Dataset is obtained from: <https://drive.google.com/file/d/1er9NJTLUA3qnRuyhfzuN0XUsoIC4a-q/view>

Dataset contains: **3,171 REAL vs. 3,164 FAKE News text. (Mostly Political Topic)**

Text Dataset is re-touched and evaluated with below for the model building:

1. Original Text	2. Lemmatized Text	3. Stemmed Text
Studies	Study	studi

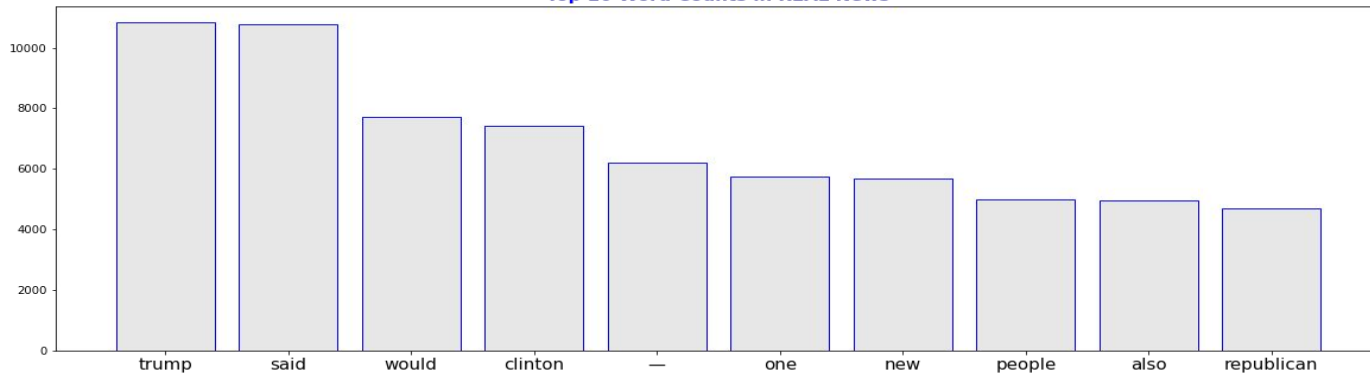
# Data Exploration



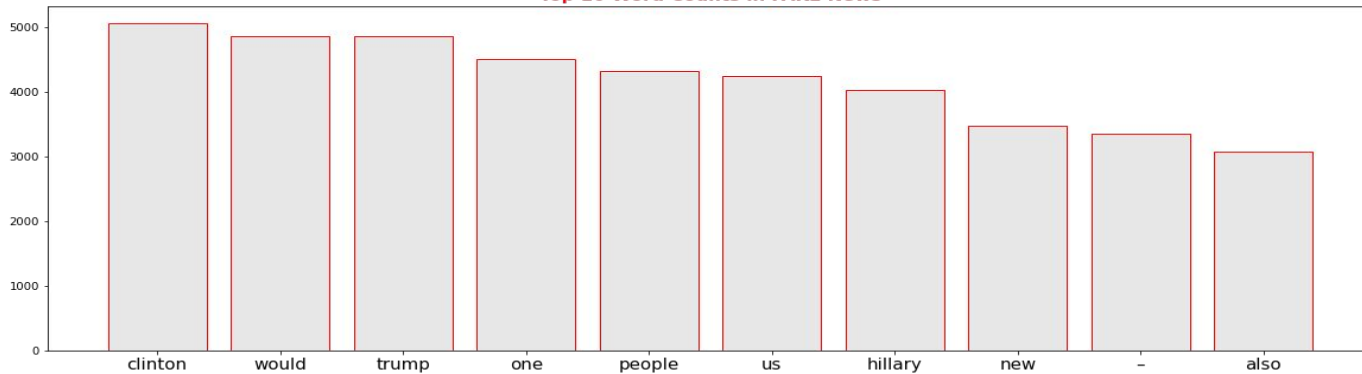
FAKE news is shorter than REAL news in terms of its length

# Data Exploration

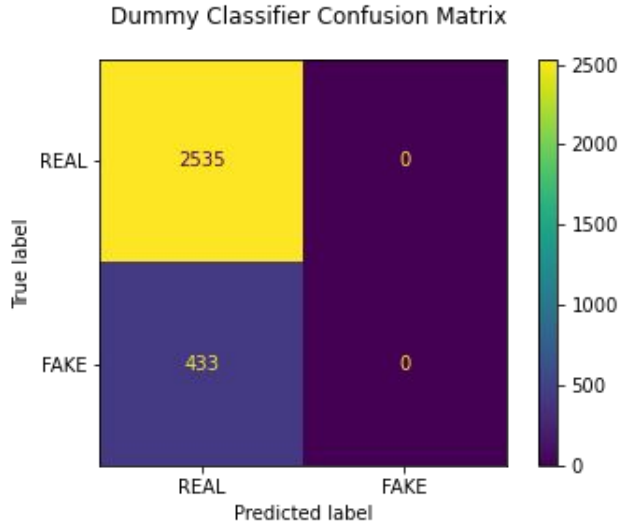
Top 10 Word Counts in REAL News



Top 10 Word Counts in FAKE News



# Model #1 - Dummy Classifier Model



- Accuracy: 0.86
- Weighted AVG F-1 Score: 0.79

Class was re-sampled with 0.17 : 1 (FAKE vs. REAL) ratio.





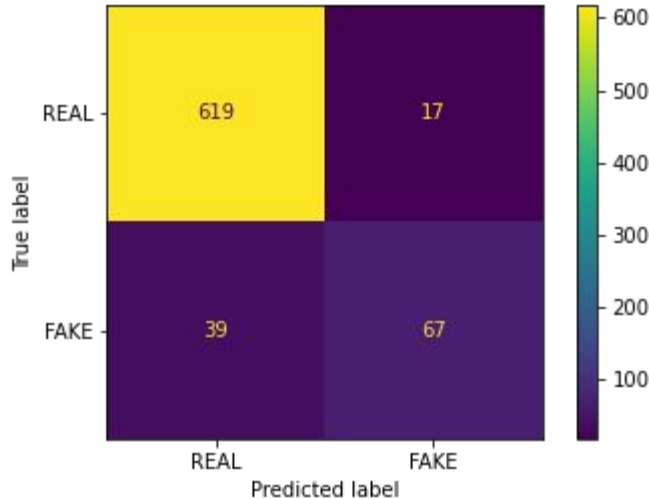
# Model Cross Validation Score Comparison

Classifier	F1 Score
Decision Tree	0.870
RandomForest	0.899
MultinomialNB	0.936
PassiveAggressive	0.937

Class was re-sampled with 0.17 : 1 (FAKE vs. REAL) ratio.

# Final Model - PassiveAggressive Classifier Model

Imbalance Data PassiveAggressive Confusion Matrix



- Accuracy: 0.92
- Precision: 0.80
- Recall: 0.63
- Weighted avg F1-Score: 0.92

TFIDF Vectorizer with the lemmatized dataset



## Next Steps

1. Try Deep Learning Models.
2. Try Web-Scraping from Fake News sites like Onions and see how well our model performs.
3. Try Web-Scraping from Social Media and see the Fake News Detection Rate.



# Thank you!

Ji Hoon Chung

LinkedIn: <https://www.linkedin.com/in/jhj1650>

GitHub: <https://github.com/jhj1650>

E-mail: [jhj1650@gmail.com](mailto:jhj1650@gmail.com)