# How to Spot a Troll Using Natural Language Processing

James M. Irving, Ph.D.

06/12/19

# Background on Russian Trolls & Internet Research Agency

- In February 2018, Robert Mueller filed indictments against Russian nationals for engaging in illegal political activities as part of his Russia investigation.
- The Internet Research Agency is a defendant, accused of being a Russian “Troll Factory” connected to 2,848 twitter handles.
- FiveThirtyEight extracted 3 million tweets from these twitter handles.

# Goals & Value for Your Organization

- Our goal is to build a machine learning model that uses Natural Language Processing alone to identity Russian Troll tweets versus authentic users.

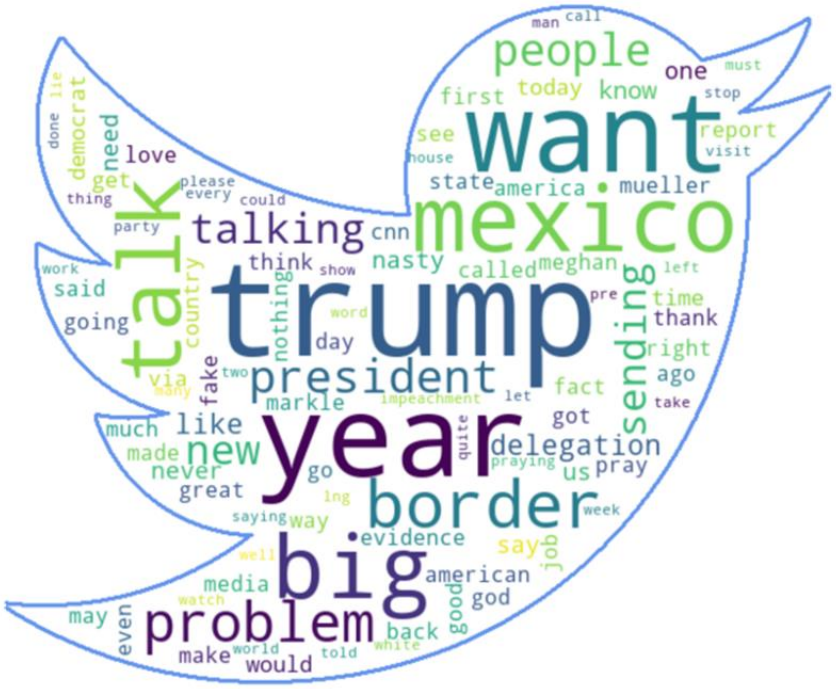
- Imagine an algorithm for your platform that can detect if user's social media messages may be foreign agents.
  - Get ahead of manipulation of your platform for malicious purposes.
  - Avoid bad press and drops in public trust/perception.

# Content Analyzed

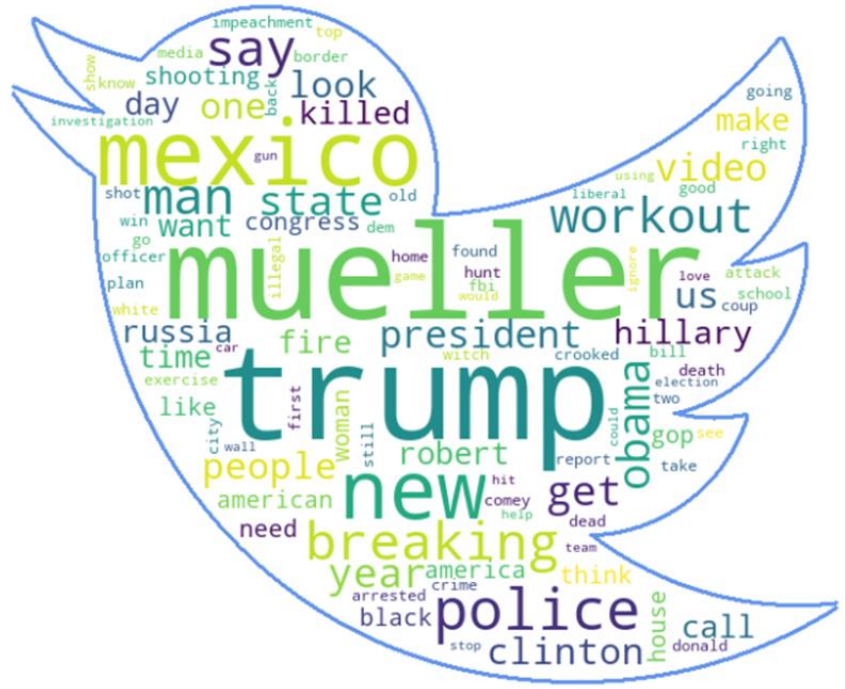
- We wanted to focus our initial analyses on the language alone.
  - To do so, we limited our analyses to only original-content tweets written in English, which totaled 1.4 millions tweets.
  - Using the Twitter API, we extracted 40,000 tweets, matching the top 40 most common @’s (mentions) and compared them to an equal-sized random sample from the 1.4 million tweets.

Most Frequently Used Words

Control

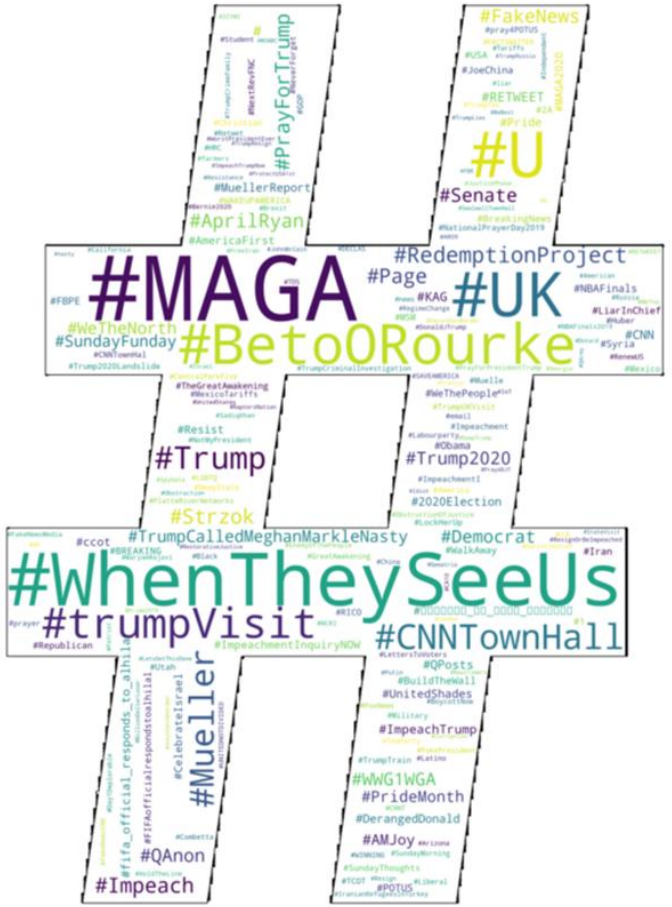



Troll



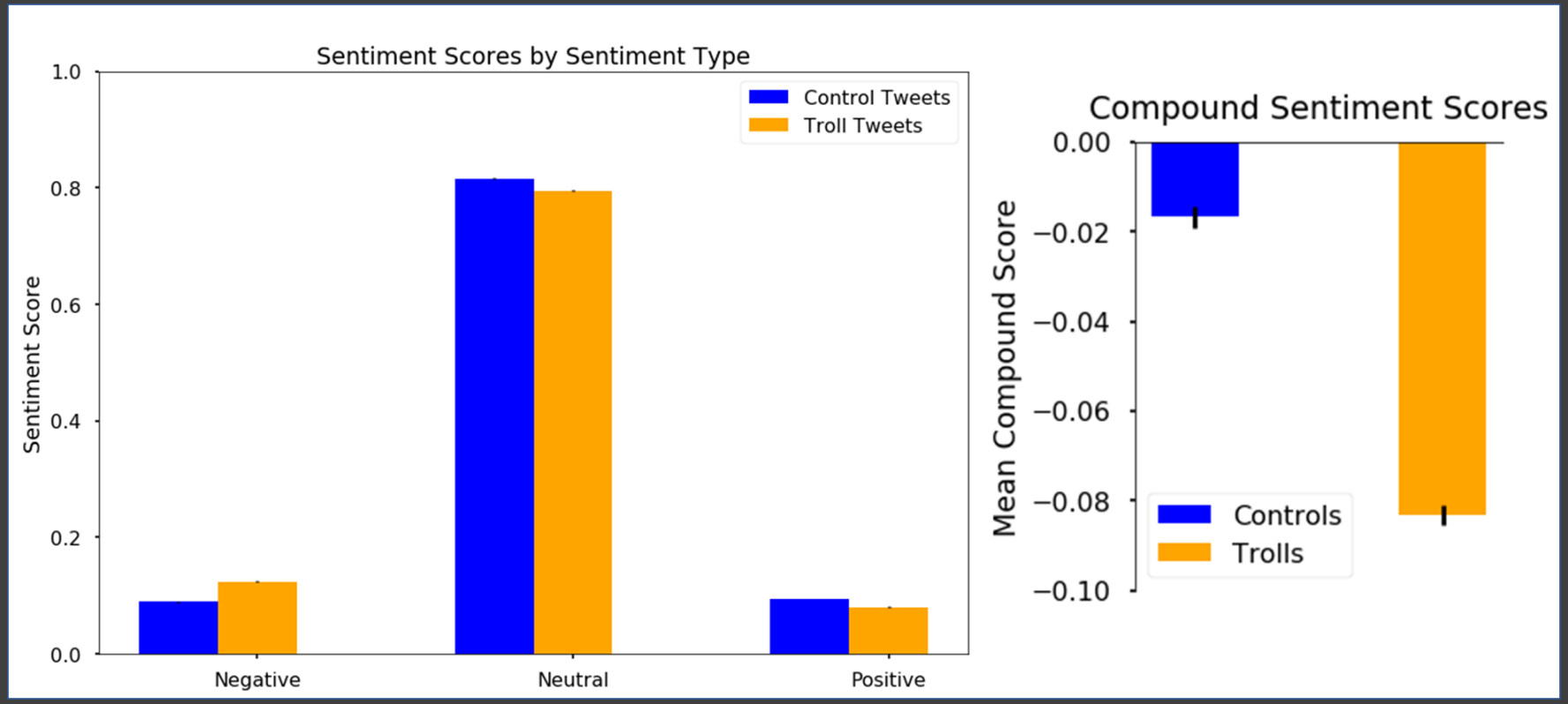
# Most Common Hashtags

## Control



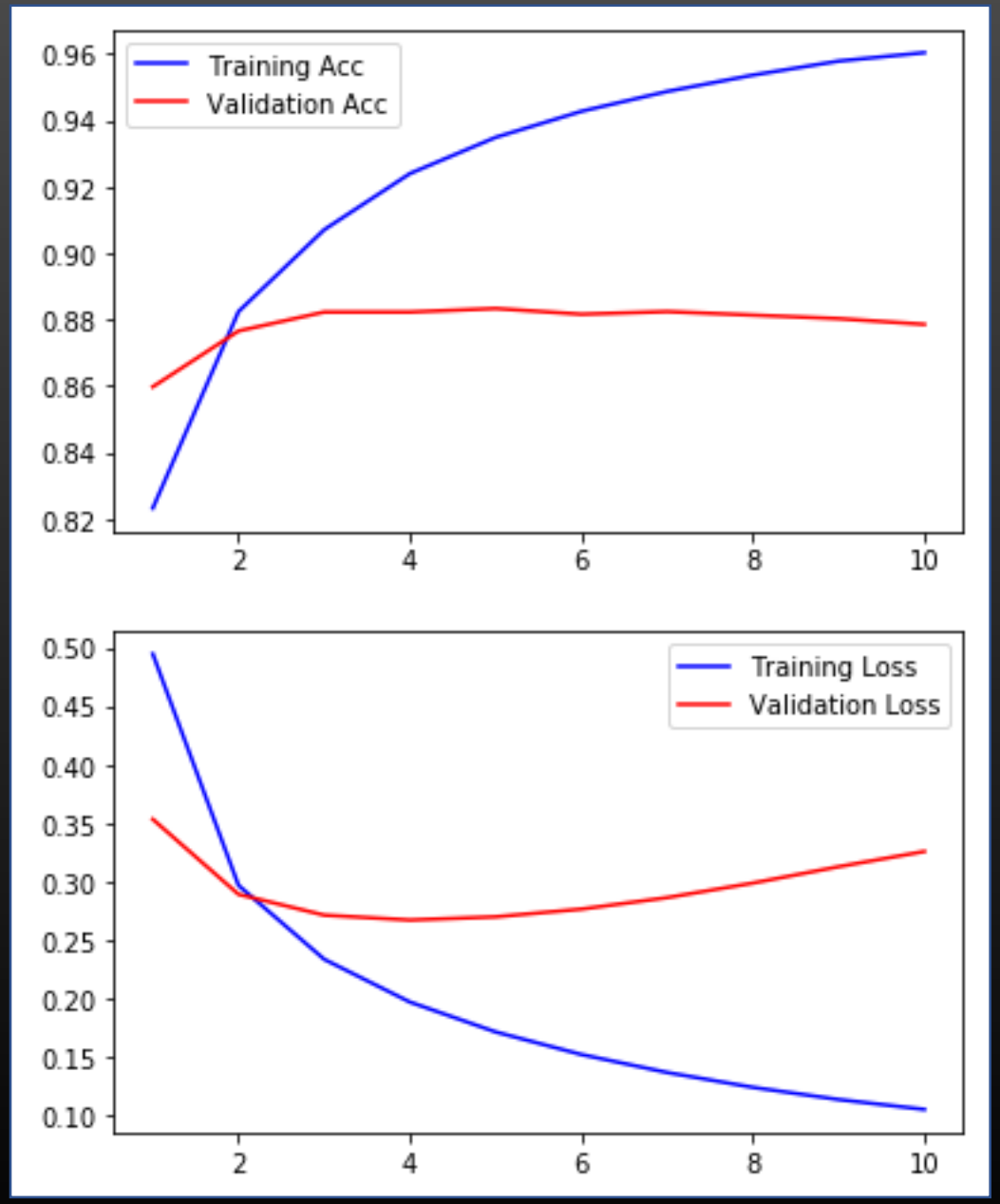

## Troll

# Sentiment Analysis



# Our Model

- Using multiple machine-learning models, including artificial neural networks, we can predict Russian troll tweets:
  - With 88-93% accuracy in 78 ms using logistic regression.
  - With 89-99% accuracy in 31 sec using artificial neural networks.



# Future Directions

- By granting our organization Enterprise-level Twitter API access, we can provide even better predictive capabilities equipped with batch historical tweets.
- We can build models upon our Natural Language Processing models to use our initial predictions combined with additional descriptive statistics.
- We can produce more sophisticated neural networks to provide better long-term analyses.

# Thank you!

- Flatiron School – Data Science Bootcamp
  - Brandon Lewis – Instructor
  - Jeff Herman – Advisor
- FiveThirtyEight – supplying troll tweet dataset