# Evolution of Hate Speech on Social Media during US Elections

Team 9: Joshua Chestnut, Shreya Kaul, Neeraj Kulkarni, Kshitij Pathak

What is the problem and what are we trying to do?

- Rise in hate speech during election on social media

- Hate Speech contributes to misinformation, formation of hate communities, and can lead to physical violence and protests

- How has hate speech on Twitter evolved during the election including post presidency?
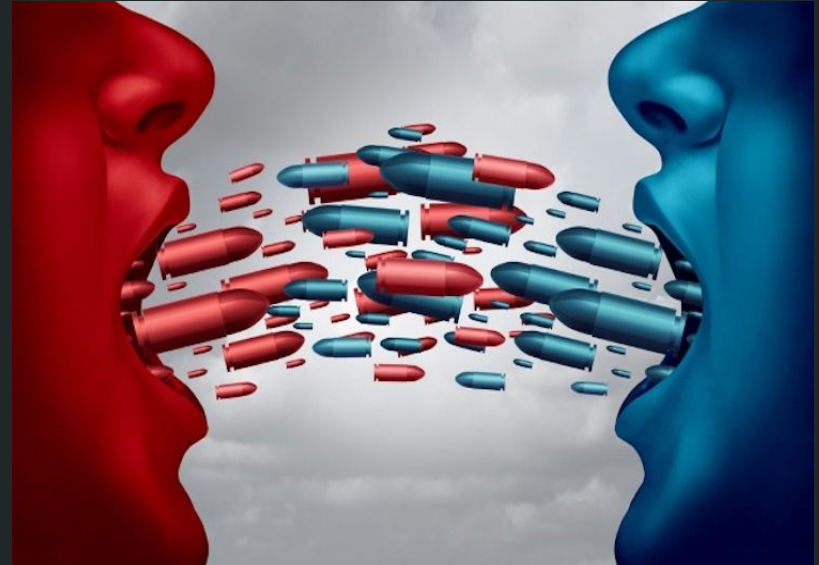
Who cares?

- Insight to how the behavior of a community evolves during election time
- Necessary to identify hate speech for content moderation and establishing social media regulations
- Social media platforms
- Government officials
- Politicians and lobbyist
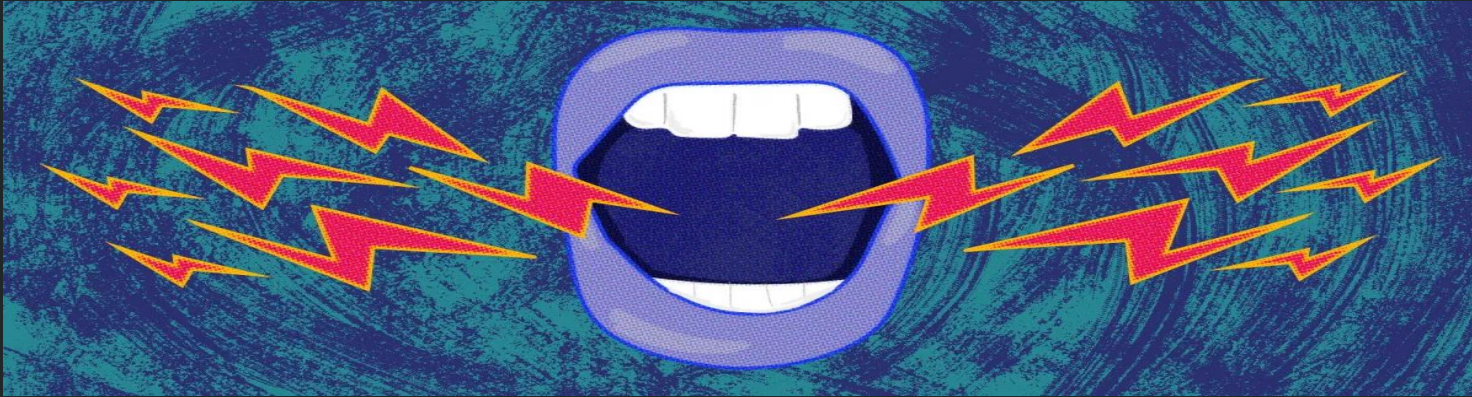
How is it done today?

- Natural Language Processing (NLP)

- Machine Learning (Training Set and Testing Set)

- Naive Bayesian

- Decision Tree

- KNN

- Logistic Regression

- Random Forest

- Confusion Matrix (To establish the accuracy of each model)

What is new in our approach and why do we think it will be successful?

- Most current research only identifies hateful speech

- We are showcasing the trends of hateful speech during the election time

- Demonstrating the proportion of hateful speech pre- and post-president inauguration

- Evaluate engagement with hate speech and non-hate speech

- If possible, we may bin users by demographics
  - This will add a layer that displays which community of users are more, or less, likely to engage in hateful speech
- Our data has already been used in other research with high accuracy. Moreover, this dataset is clean and has enough sample set to create a good classification model.

- A high usability index of this dataset makes this project likely to succeed.

How will it be measured?

- Proportion of hateful tweets pre- and post-election.

- Possible Method 1:

$$\frac{\text{No. of tweets from selected keywords that contain hate speech}}{\text{All tweets from selected keywords}}$$

- Possible Method 2:

Establish an engagement criteria using attributes like 'no. of comments', 'no. of retweets', etc., to compare and contrast hate speech and non hate speech tweets

What difference will it make?

- Government officials can take preventative measures against those communities that engage in hate speech

- Hateful speech online can lead to protest and physical violence

- Across time, what are the demographics of individuals who engage in hate speech?

- Potential for social media platforms to use this information to push a particular political agenda

- Allows social media platforms to establish necessary regulations -- particularly concerning the demographics of the communities and what type of speech is most engaged with online

Data Sources

- Hate Speech and Offensive Language Dataset (https://www.kaggle.com/mrmorj/hate-speech-and-offensive-language-dataset)
  - CCO: Public Domain
  - This dataset will be used to train the model

- Twitter API
  - Tweets collected one month pre- and post-presidential inauguration
  - These tweets will pertain to United State politics (e.g. #politics, #president, #election, etc.)
  - Geographic filter for United States
  - These tweets will compose the testing set

Why is it hard?

- Training an accurate model - will it perform as well with the data we collect?

- Collecting appropriate twitter data - sample data must be representative of the population

- Managing the volume of tweets based on appropriate keywords and filtering

- Rate limit

- Filtering by specific time period and demographics (age, location, etc.)

Plan of Activities

Week	Feb 8-12	Feb 15-19	Feb 22-26	March 1-5	March 8-12
Phase	Project Pitch Due Feb 8th	Project Phase 1 Due Feb 22nd	Project Phase 2 Due March 1st	Project Phase 3 Due March 11th	Project Phase 3 Due March 11th
Activity	Planning the project and Twitter data collection Determine methodology	EDA and preliminary analysis, identifying any possible roadblocks	Processing, visualizing, and analyzing data Selecting high accuracy ML model based on experimentations	Executing trained ML model on twitter corpus Interpreting and reporting findings	Wrap-up / Finalize Report
Team Meeting	2/10/2021	2/17/2021	2/24/2021	3/3/2021	3/10/2021