

#AnalyzingHate SpeechOnTwitter

@ChristopherDeLaCruz (not my actual twitter
handle)



#WhyHateSpeech?

- Hate crimes & speech have been on a steady rise since 2014
- Understanding the patterns behind recent hate speech trends will help us better understand which classes are at highest risk
- Precision is paramount, recall is secondary



#WhatIsHateSpeech?

In simplest terms, hate speech is derogatory language aimed at a person based on an identification class (race, gender, class, religion, sexual orientation, etc)



#HateSpeechData

Approximately 70,000 labeled tweets (Hate tweets and Not Hate Tweets) were gathered from four different Kaggle datasets and one dataset from Georgia Tech's CLAWS

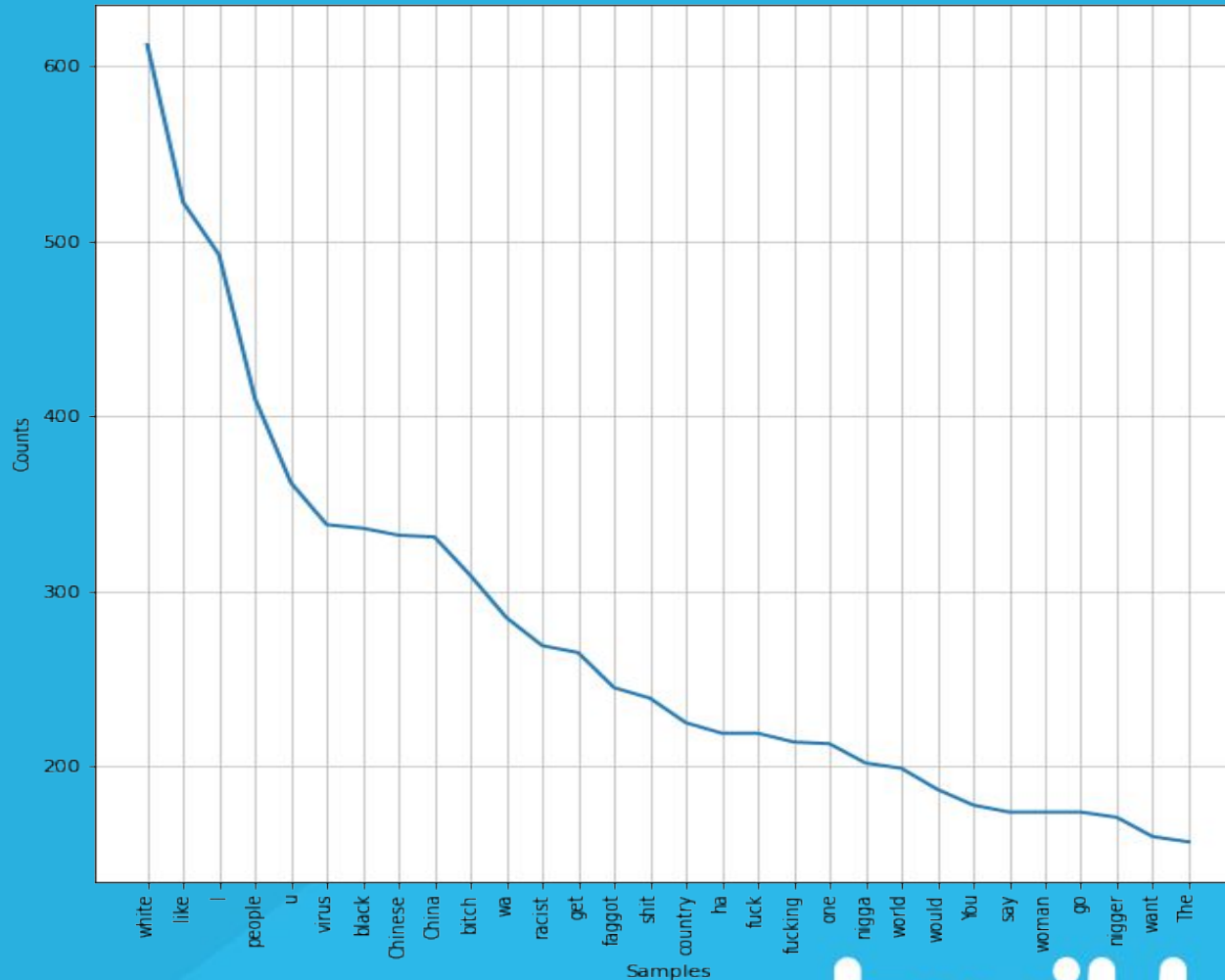
Labels are:

1 - Hate Speech

0 - Not Hate Speech

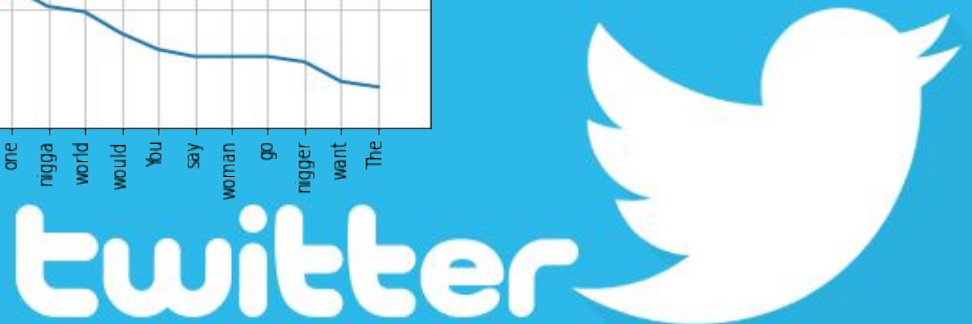


#HateSpeechResults

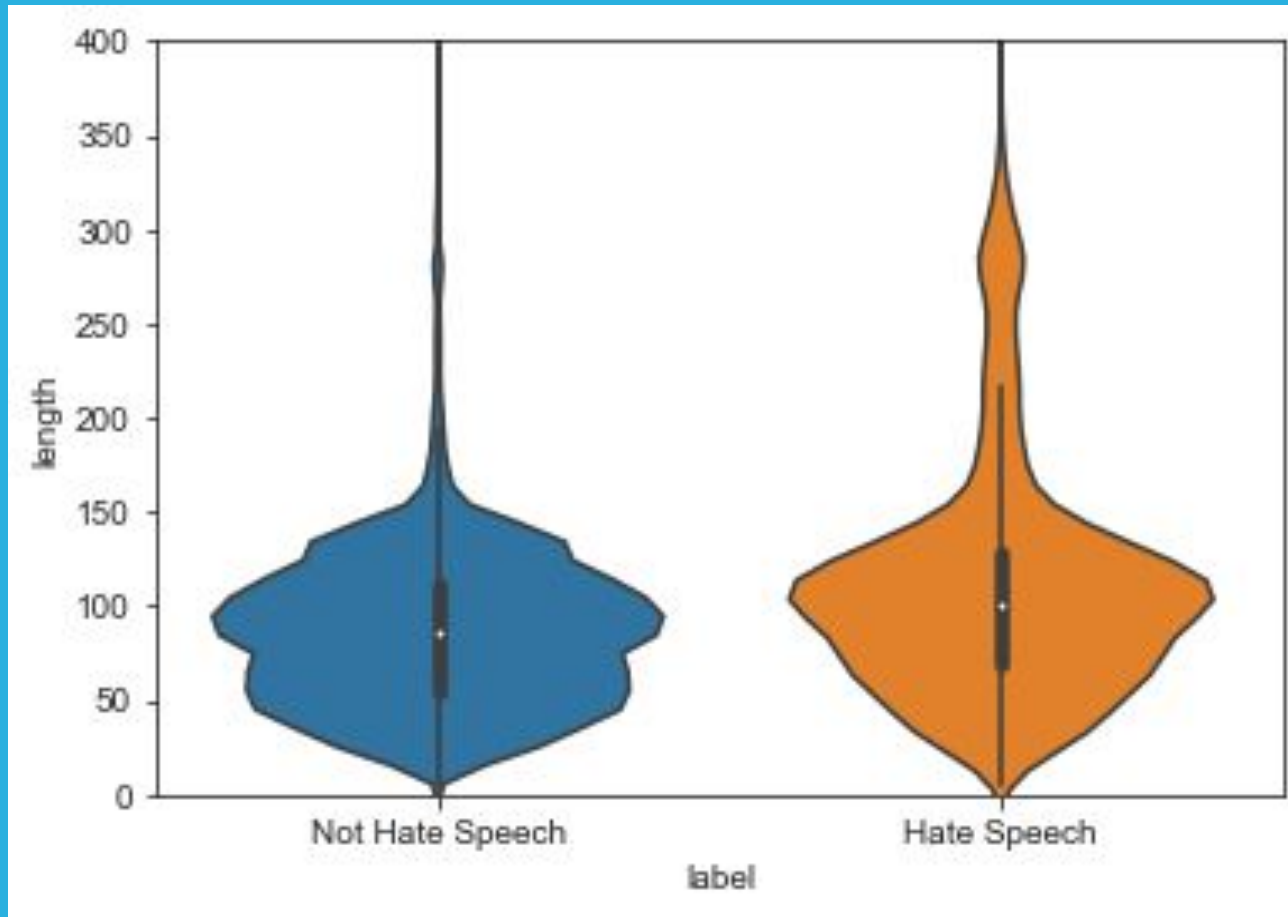


Identity Classes Mentioned Most Often:

- Race: Asian
 - Race: Black
 - Race: White
- Gender: Women
- Sexual Orientation: Gay



#HateSpeechResults



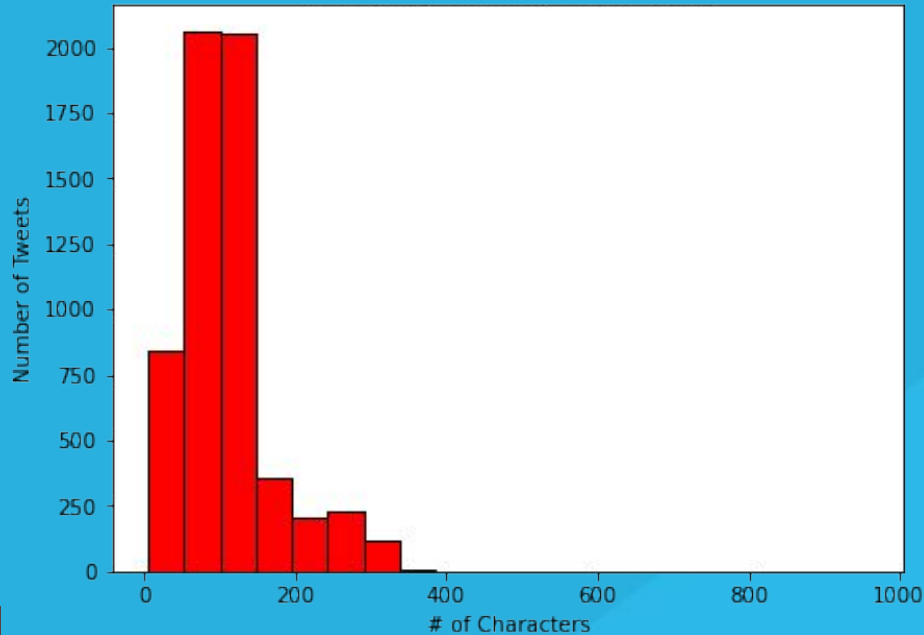
Hate tweets
look longer
and the math
agrees but
how much
longer?

twitter

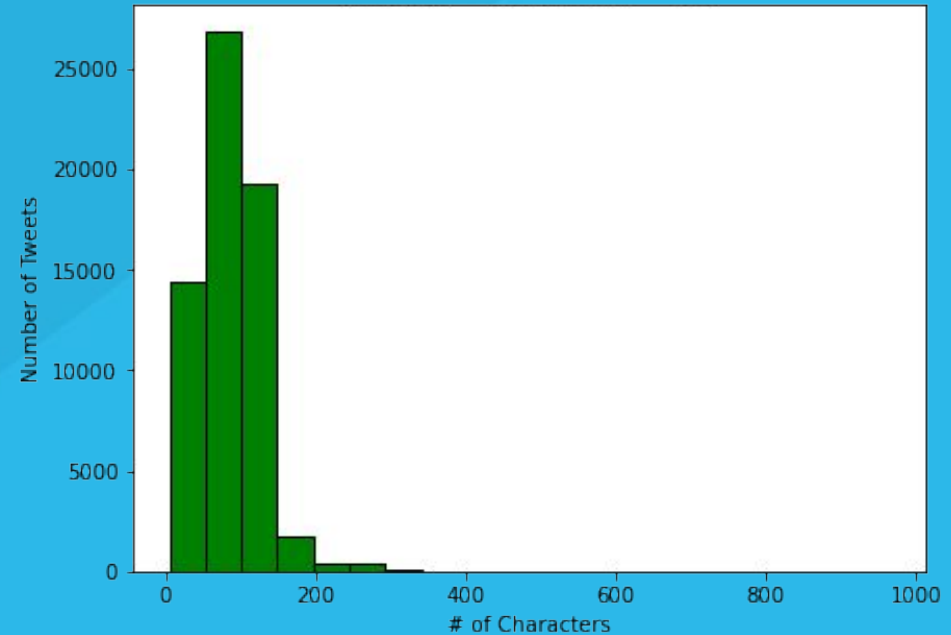


#HateSpeechResults

Character Distribution of Hate Tweets / mean = 109.92
/ median = 101.0 / mode = 106



Character Distribution of Not Hate Tweets / mean = 86.3
/ median = 86.0 / mode = 109



Hate tweets are about 24 characters longer on average



#HateSpeechResults

Over 7,000 models were run but there could only be ONE winner.

Ultimately the best model was...



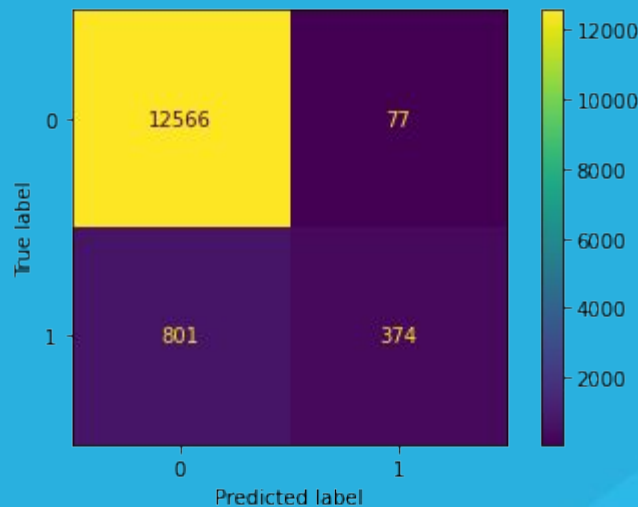
#DRUMROLL

like a literal drumroll, this is not the name of the model



#HateSpeechResults

Support Vector Classifier!



- Out of the 451 tweets that the model predicted as hate speech, 374 tweets were actually hate speech (83% precision)
- Out of 1,175 hate tweets, the model correctly identified only 374 of these tweets (32% recall)



#HateSpeechResults

The model excels at identifying hate speech revolving around current events (pandemic, politics, etc).

Trending Hashtags on Hate Tweets Model Got Correct

- #allahsoil (regarding a book titled 'Allah's Oil')
- #politics
- #trump
- #libtard
- #liberal
- #chinesevirus



#NextSteps

- We need more tweets. A LOT more tweets.
- Look into ways to continue improving the model's ability to detect hate speech (more data, different sampling techniques, etc)



#ThankYou!

Presentation By Christopher de la Cruz

Email: cdelacruz2013@gmail.com

Github: cdlc01

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

