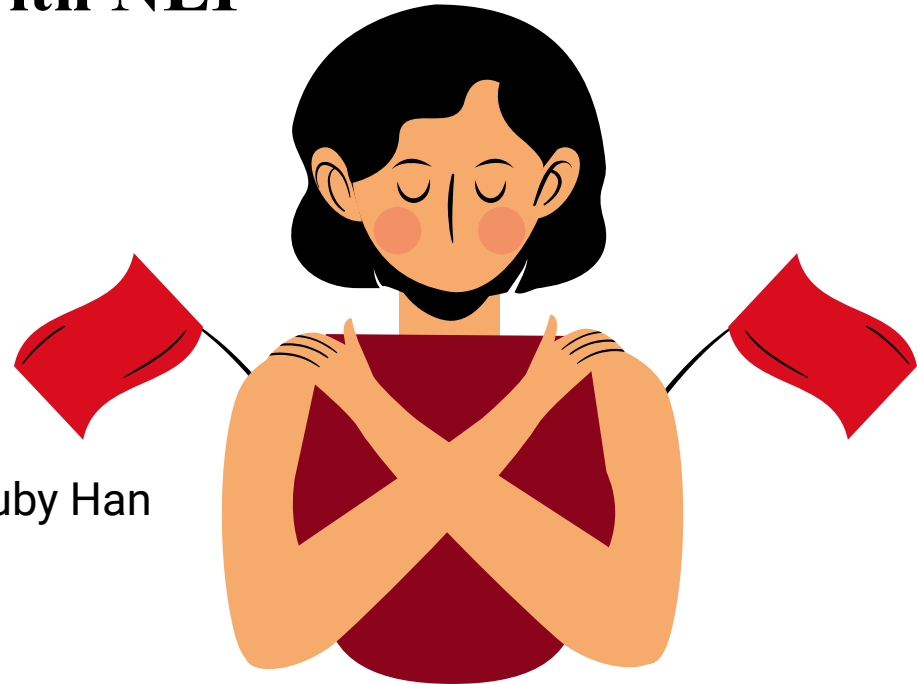# Increasing Efficacy of Federal **Firearm** Background Checks by Identifying Patterns of **Threatening** Language with NLP

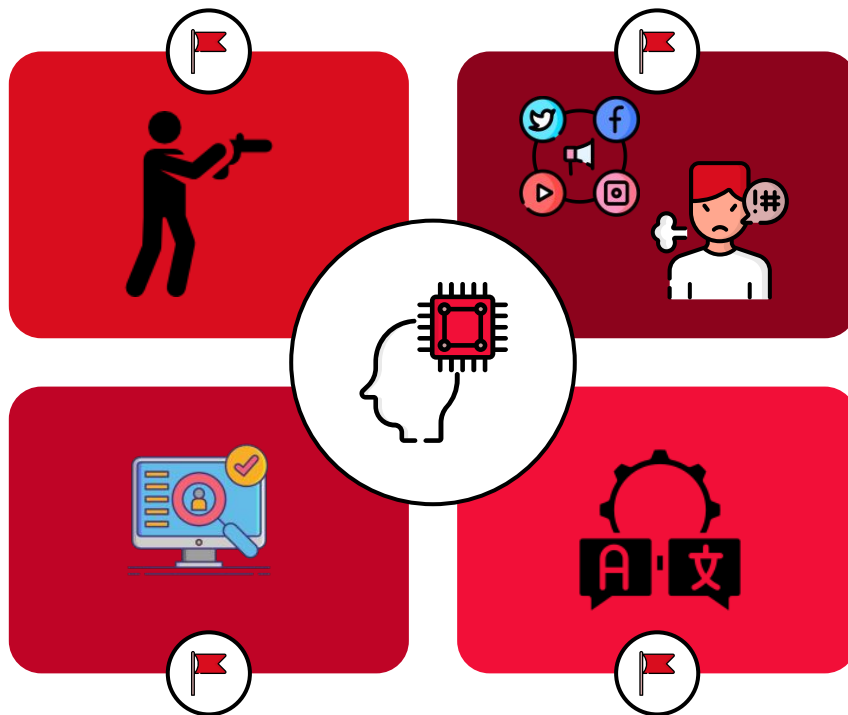Presented By: Aditya Bajaj, Gerrit Lensink, Ruby Han
Date: August 4th, 2022

Introduction

Mass Shooting Events

Recent rise in firearm tragedies across United States

Current Firearm Regulations & Background Checks

- Criminal history
- Mental health history
- Civil cases (domestic violence)

Social Media Scraping

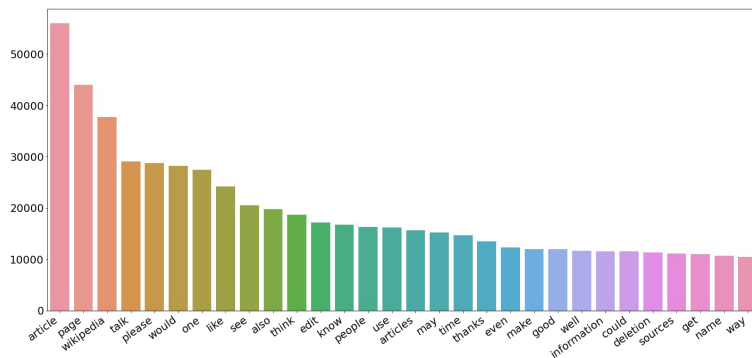Make use of publicly available comments

NLP Techniques

Classification model which identifies user-generated text as violent

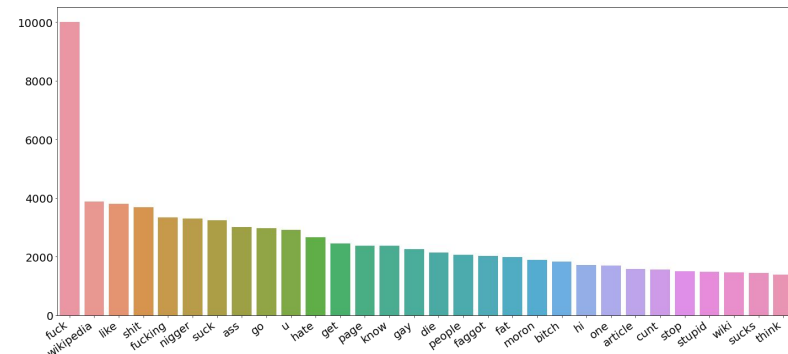# EDA - Top Words

Words frequented in Non-Violent Comments

Words frequented in Violent Comments

Top Words in Non-Violent Comments

Top Words in Violent Comments

Transforms

Cosine Similarity

See next slide for example

Upper Case + Exclamation + Question Mark

- 14% vs 5%
- 3.47 vs 0.34
- 0.59 vs 0.43

Special Characters

@ - … $ { } \n \t etc.

Masking Profanity
Used *better-profanity* to identify profanity & mask words

Summarize Comments

Used BART in attempt to fix comment length
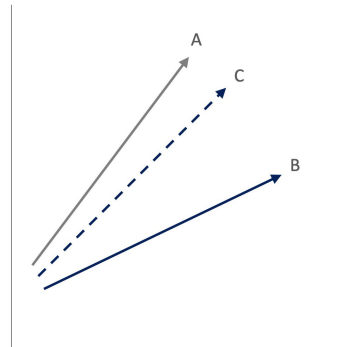
Remove Sensitive Information

- IP addresses
- Usernames

Embedding Coercion

Sentence Text

A	To kill a mockingbird is a pretty sad book that even deals with murder
B	I am going to FIND YOU and KILL YOU!
C	I am going to find you and kill you.

Cosine Similarity

	A	B	C
A	1		
B	0.76	1	
C	0.8	0.84	1

BART Comment Summary

Original Comment:
["   interesting page   There's some good info here, but as Calum says, there "
 'are also some errors here. By the Right Quick March The Pipe Majordirects '
 'the band with words that express the tempo   Right Quick March. A right '
 "quick march doesn't exist. The Drum Major directs the band to play a quick "
 'march, yes, but the phrase by the right is not related to the tempo. This '
 'tells the band which file in this case, the right file to take their '
 'dressing from. The right is usually the side the pipe major is on, either '
 'on his own, or as the right hand person in the front rank. You often hear '
 'by the centre as well, with by the left being far less common.']

BART Comment Summary:
['The Pipe Majordirects the band with words that express their tempo. The '
 "right quick march doesn't exist, it's just a phrase to be used when you're "
 'marching in front of your drum major or other pipemaster']
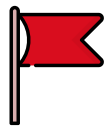
Average Violent Comment Length: 300
Average Non-Violent Comment Length: 399

Median Violent Comment Length: 127
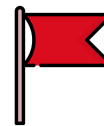Median Non-Violent Comment Length: 213

Model Metrics

F1 Score

Precision

- Metric to reduce number of false positives
- Low due to high false positive counts

Recall

- Metric to reduce number of false negatives
- High due to low false negative observations

Results

Model	Description	F1 Score
Baseline BERT	No transformations	**64.9%**
Interim BERT A	Unsuccessful transforms	45.1%
Interim BERT B	12 hidden layers	67.6%
BERT + CNN	Added CNN output layer	69.2%
DistilBERT (final)	All transforms included	**70.1%**

Final Model Architecture: DistilBERT

Hyperparameter	Value
Sequence Length	128
Train/Val Split	80/20
Batch Size	32
Epochs	6
Dropout	0.1
Learning Rate	1e-5
Optimizer	AdamW
Hidden Layer Size	50
Hidden Layer Count	2
Undersampling	None

Questions?