Multi-Modal Classification for Polarization Intent Detection in Social Media

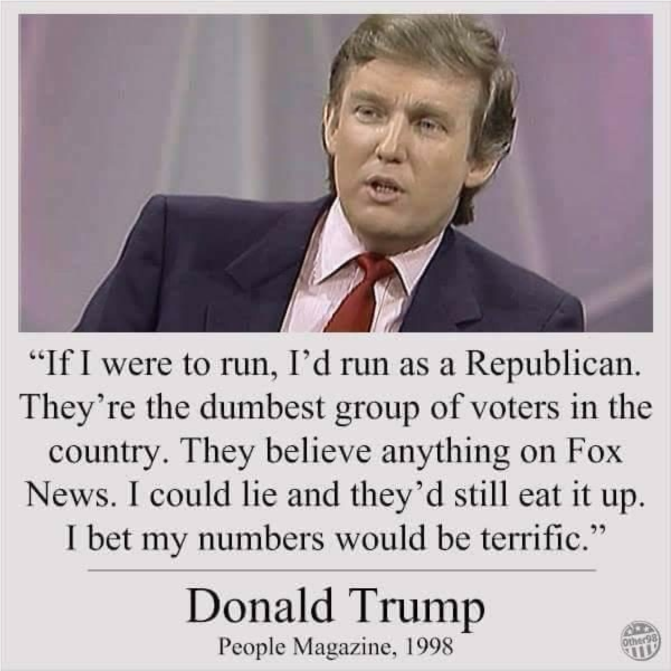
Tobby Lie, Haadi Jafarian, Stephen Hartnett, Hamilton Bean, and **Farnoush Banaei-Kashani**

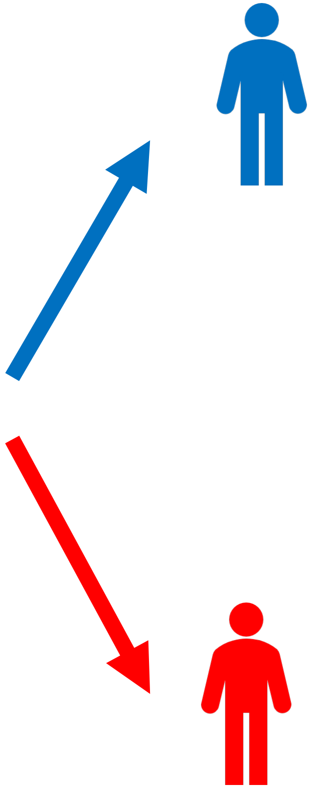
Contact: farnoush.banaei-kashani@ucdenver.edu

CU Denver Data Science Symposium 2021
August 27, 2021

Motivation

- We are motivated by the study of a single agent/group with the intention to polarize a network of social media users using multi-modal memes.

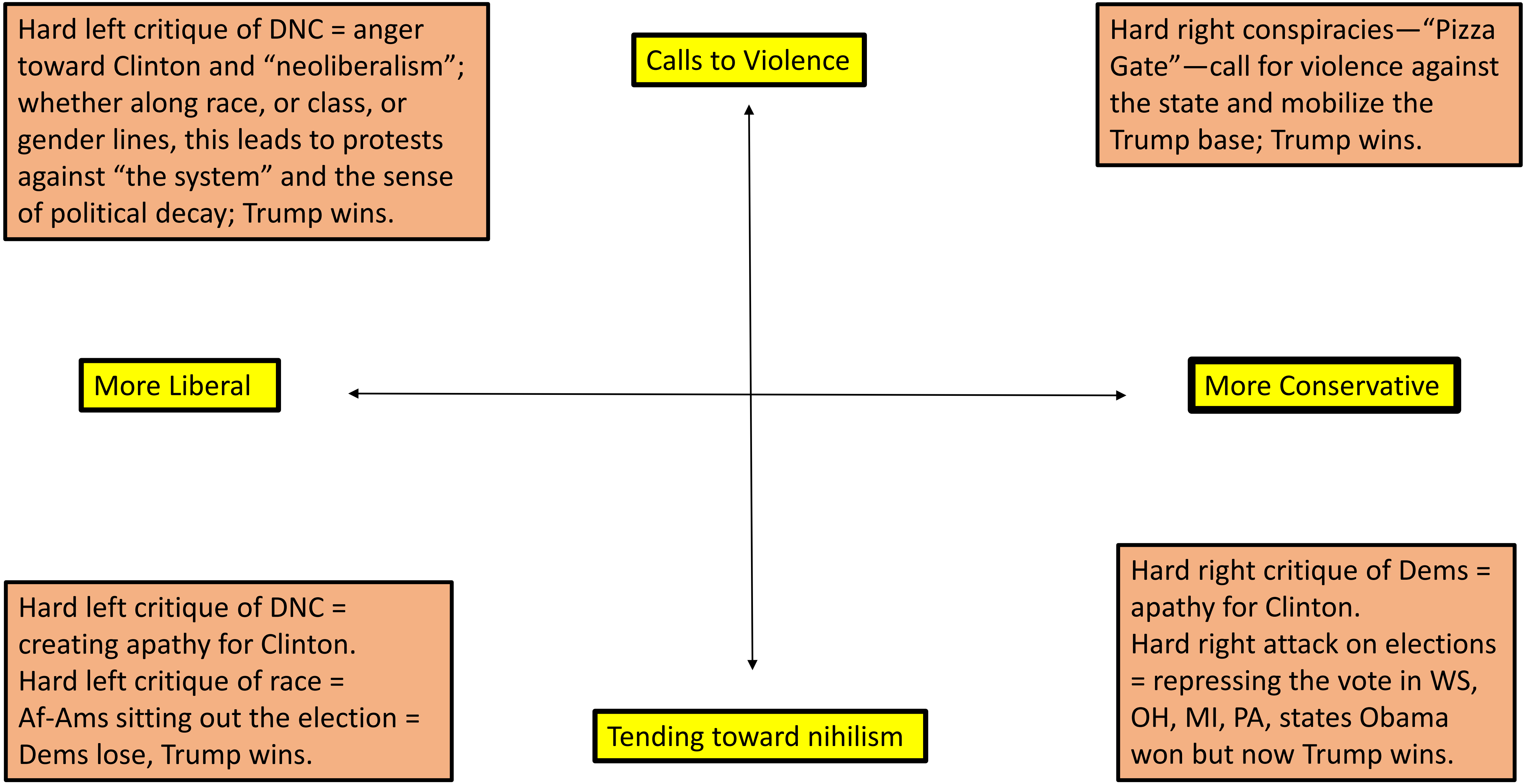




Problem Definition

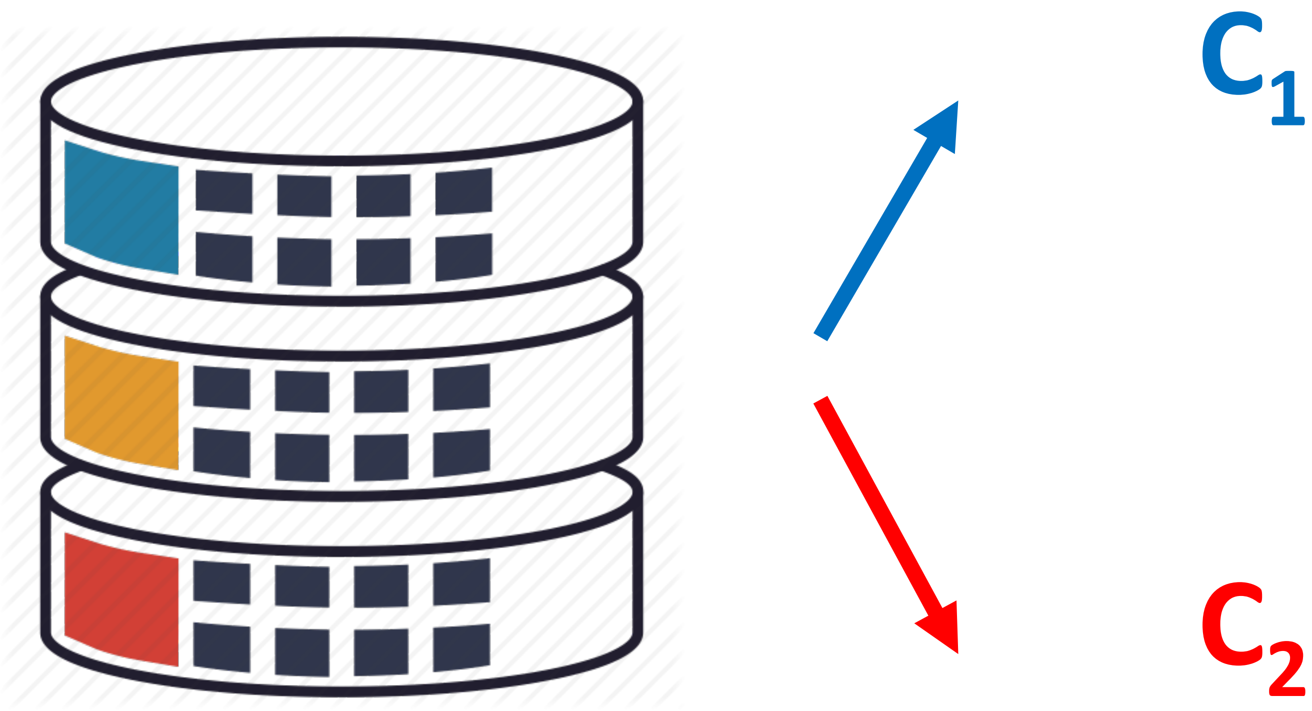
Given a corpus of memes posted by a single entity on social media, we seek to detect the intention of polarizing.

Map Courtesy of Prof. Stephen Hartnett



Problem Definition

Given a corpus of memes posted by a single entity on social media, we seek to detect the intention of polarizing.



Dataset

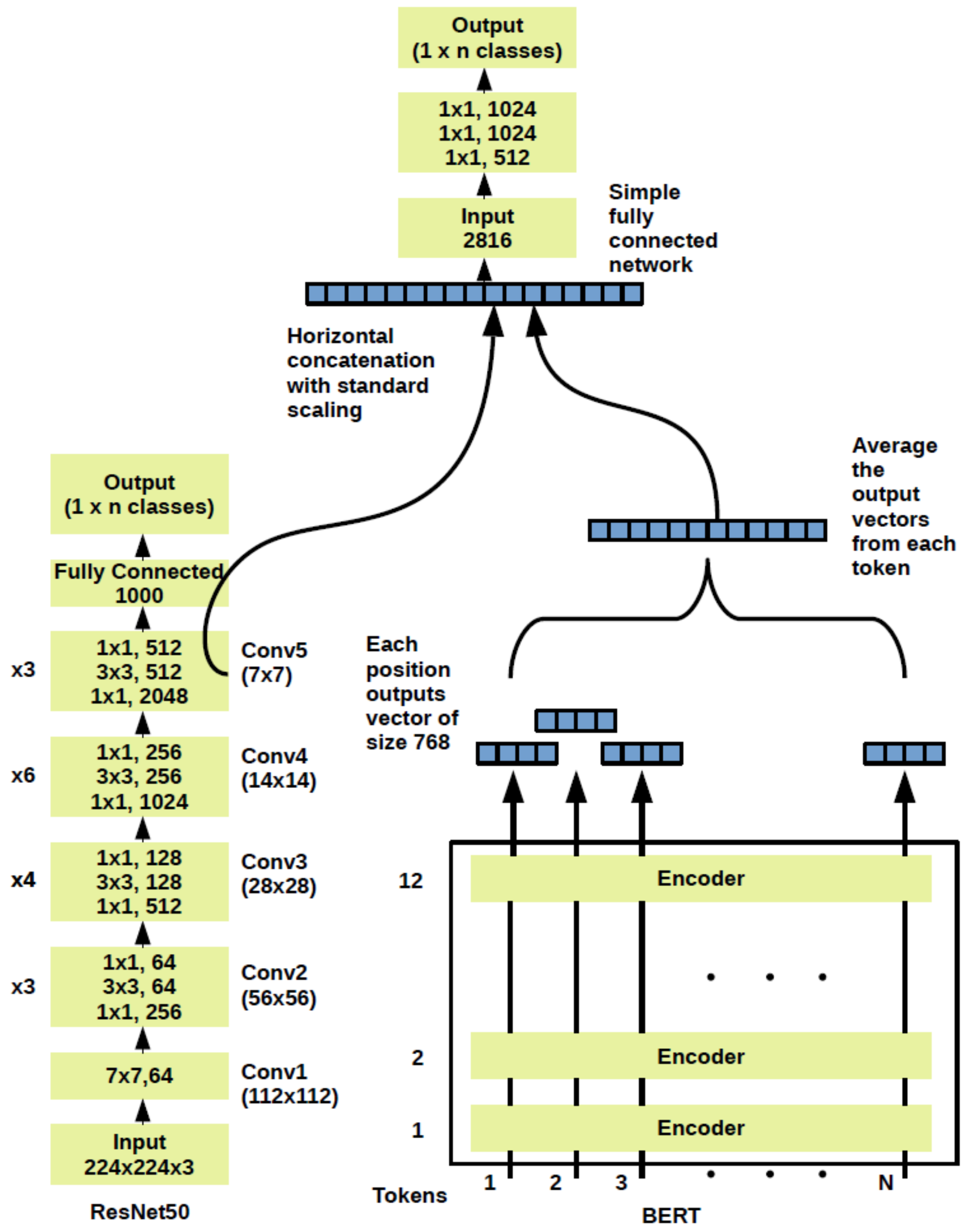
University of Maryland College Park curated dataset

- Internet Research Agency (IRA) data
- 3012 memes

- Notable metadata information:
  - URL to the original ad
  - Identifier
  - Title
  - **Description text**
  - **Image URL**
  - Tags
- “anger” and “fear” tags formed a basis for our class labels: “Violence” and “Apathy”
- Class imbalance issue

Integrated Multi-Modal Model

- Single feature vector required from separate text and image embeddings
- Image embedding through trained ResNet50
- Text embedding using BERT language model
- Simple horizontal concatenation with normalization fusion strategy
- Fused vector fed into simple feed-forward network



Experimental Results

Support Vector Machine	Acc.	Prec.	Rec.	F1
Imbalanced Violence	0.74	0.45	0.51	0.48
Undersampled Violence	0.69	0.64	0.72	0.68
Oversampled Violence	0.80	0.85	0.79	0.82
Imbalanced Apathy	0.88	0.32	0.38	0.35
Undersampled Apathy	0.66	0.65	0.63	0.64
Oversampled Apathy	0.91	0.95	0.87	0.91

SmallerVGGNet				
Imbalanced Violence	0.67	0.78	0.75	0.77
Adjusted Class Weights Violence	0.56	0.94	0.0.43	0.59
Undersampled Violence	0.56	0.71	0.27	0.39
Oversampled Violence	0.75	0.73	0.77	0.75
Imbalanced Apathy	0.88	0.88	**1.0**	0.94
Adjusted Class Weights Apathy	0.88	0.88	**1.0**	0.94
Undersampled Apathy	0.58	0.58	0.65	0.61
Oversampled Apathy	0.64	0.63	0.72	0.67

Naive Bayes				
Imbalanced Violence	0.77	0.54	0.58	0.56
Undersampled Violence	0.73	0.82	0.70	0.76
Oversampled Violence	0.77	**0.89**	0.73	0.80
Imbalanced Apathy	0.90	0.23	0.55	0.32
Undersampled Apathy	0.73	0.76	0.69	0.72
Oversampled Apathy	0.85	0.96	0.79	0.87

ResNet50				
Imbalanced Violence	0.76	0.79	0.90	0.84
Undersampled Violence	0.60	0.77	0.35	0.48
Oversampled Violence	0.87	0.87	0.87	0.87
Imbalanced Apathy	0.93	0.99	0.93	0.96
Undersampled Apathy	0.92	0.95	0.89	0.92
Oversampled Apathy	0.95	**0.99**	0.92	0.95

Multi-Modal Fusion				
Imbalanced Violence	0.86	0.76	0.70	0.73
Undersampled Violence	0.85	0.82	0.89	0.85
Oversampled Violence	**0.91**	0.88	**0.95**	**0.91**
Imbalanced Apathy	0.92	0.57	0.63	0.60
Undersampled Apathy	0.82	0.77	0.89	0.82
Oversampled Apathy	**0.97**	0.94	**1.0**	**0.97**

Conclusions and Future Work

- We present a scalable and automated implementation of polarization intent detection
- The proposed model can be utilized via transfer learning for similar intent detection tasks.

- Future Directions
  - CoronaVirusFacts Alliance: https://www.poynter.org/coronavirusfactsalliance/
  - GeoCoV19: https://crisisnlp.qcri.org/covid19