

# Remote Canvassing

Suleyman Qayum

# INTRODUCTION

## ❖ What is *Remote Canvassing*?

- *predict a person's stance on one or more political issues*
- *use these predictions to identify potential voters/supporters*

# OUTLINE



BUSINESS  
CONTEXT



DATA



MODEL



RESULTS



EVALUATION



LIMITATIONS &  
NEXT STEPS

# BUSINESS CONTEXT

- ❖ Traditional canvassing is inefficient
- ❖ Make this process more efficient by utilizing **online forums/discussions**
  - *less time consuming*
  - *greater reach*

# DATA

❖ Comments scraped from *Reddit*

❖ Represent *Conservative/Liberal* stance on:

- *Abortion*
- *Immigration*
- *Healthcare*
- *Gun Control*
- *Climate Change*

# MODEL

- ❖ Receives some text input

- *can be attributed to an individual*
- *discusses/conveys political views*

- ❖ Analyzes text and determines:

- ***Issue*** → *Abortion / Healthcare / Gun Control / Immigration / Climate Change*
- ***Stance*** → *Conservative / Liberal*



# RESULTS

---

# Predicting Stance

❖ Out of all the comments that took a **Conservative** stance:

- 81% were ***correctly labeled*** as Conservative
- 19% were ***incorrectly labeled*** as Liberal

❖ Out of all the comments that took a **Liberal** stance:

- 48% were ***correctly labeled*** as Liberal
- 52% were ***incorrectly labeled*** as Conservative





# Predicting Issue

- ❖ Out of all the comments discussing **Abortion**:
  - *77% were correctly identified as such*
- ❖ Out of all the comments discussing **Climate Change**:
  - *70% were correctly identified as such*
- ❖ Out of all the comments discussing **Gun Control**:
  - *65% were correctly identified as such*
- ❖ Out of all the comments discussing **Healthcare**:
  - *70% were correctly identified as such*
- ❖ Out of all the comments discussing **Immigration**:
  - *60% were correctly identified as such*

# EVALUATION

- ❖ Model did not generalize well when given new data
  - ❖ 10% *less accurate predicting Stance*
  - ❖ 9% *less accurate predicting Issue*
- ❖ Performance shows promise
  - *model learned from ~43,000 comments*
    - *relatively small amount of data*
- ❖ Performance differed between labels

# LIMITATIONS & NEXT STEPS

## ❖ Insufficient amount of data

- *multitude of labels means a lot of data is needed*
- *more comments need to be collected*
- *need to identify more Subreddits*

## ❖ Imbalanced dataset

- *not all labels represented equally*
- *model struggles with labels that are under-represented*

## ❖ Issue selection

- *selection of issues is not exhaustive*
- *some issues overlap with others*



# THANK YOU

**Email:** [sqayum33@gmail.com](mailto:sqayum33@gmail.com)

**Github:** [@sqayum](https://github.com/sqayum)

