

# VoteMatch

Exploring Voter Profiles and Party Preference Using CES2021



**GLOCAL**  
FOUNDATION OF CANADA

Ran Shi

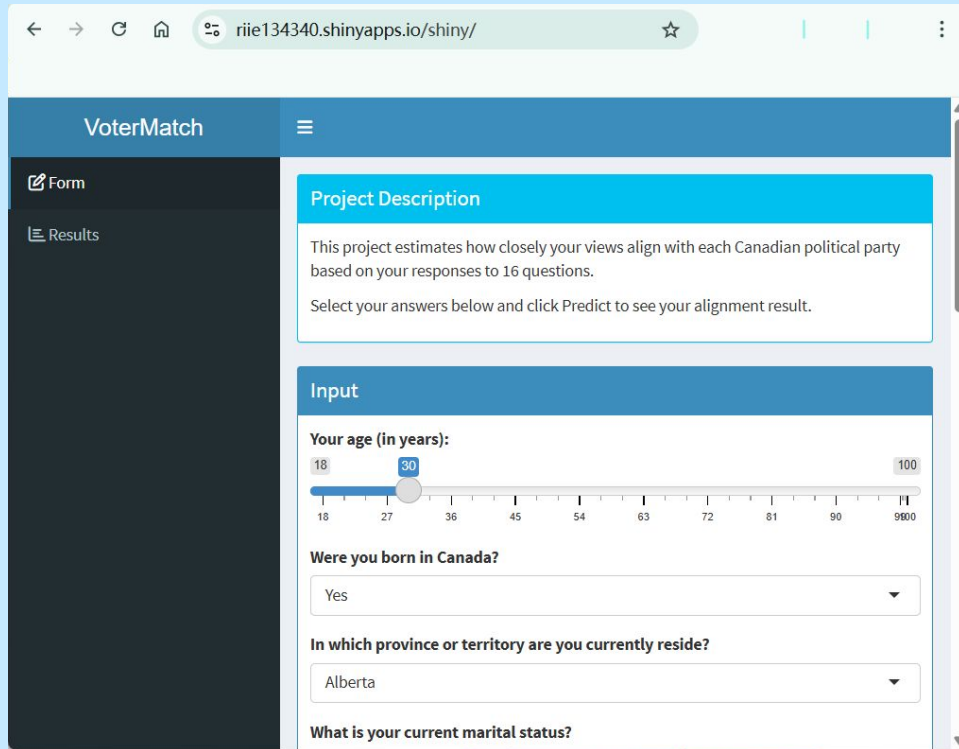
CSJ Data Analytical Specialist



# Introduction

- This tool estimates how closely an individual's views align with Canadian political parties.
- **Motivation:** make politics more approachable for people who are less familiar or less interested
- **How:** showing how people with similar characteristics have voted in past elections
- **Dataset:** 2021 Canadian Election Study (CES)
- <https://riie134340.shinyapps.io/shiny/>

# How the Tool Works



The screenshot shows a web browser window with the URL `riie134340.shinyapps.io/shiny/`. The application is titled "VoterMatch" and has a sidebar with "Form" and "Results" options. The main content area is divided into two sections: "Project Description" and "Input".

**Project Description:**

This project estimates how closely your views align with each Canadian political party based on your responses to 16 questions.

Select your answers below and click Predict to see your alignment result.

**Input:**

Your age (in years):

18 30 100

18 27 36 45 54 63 72 81 90 9600

Were you born in Canada?

Yes

In which province or territory are you currently reside?

Alberta

What is your current marital status?

- **Form Page:** complete 16 questions about demographics and attitudes
- **Prediction:** model estimates your alignment with each major Canadian party
- **Results Page:** shows how people with similar profiles voted, with probabilities for each party

## Form Page

1:45

VoterMatch

### Project Description

This project estimates how closely your views align with each Canadian political party based on your responses to 16 questions.

Select your answers below and click Predict to see your alignment result.

### Input

**Your age (in years):**

18

30

100

18 27 36 45 54 63 72 81 90 99 100

**Were you born in Canada?**

Yes

**In which province or territory are you currently reside?**

Alberta

Strongly disagree

**Too many recent immigrants are not trying to integrate into Canadian society.**

Strongly disagree

**Immigrants take jobs away from other Canadians.**

Strongly disagree

**Do you think income inequality is a serious problem in Canada?**

Strongly disagree

**The government should leave it entirely to the private sector to create jobs.**

Strongly disagree

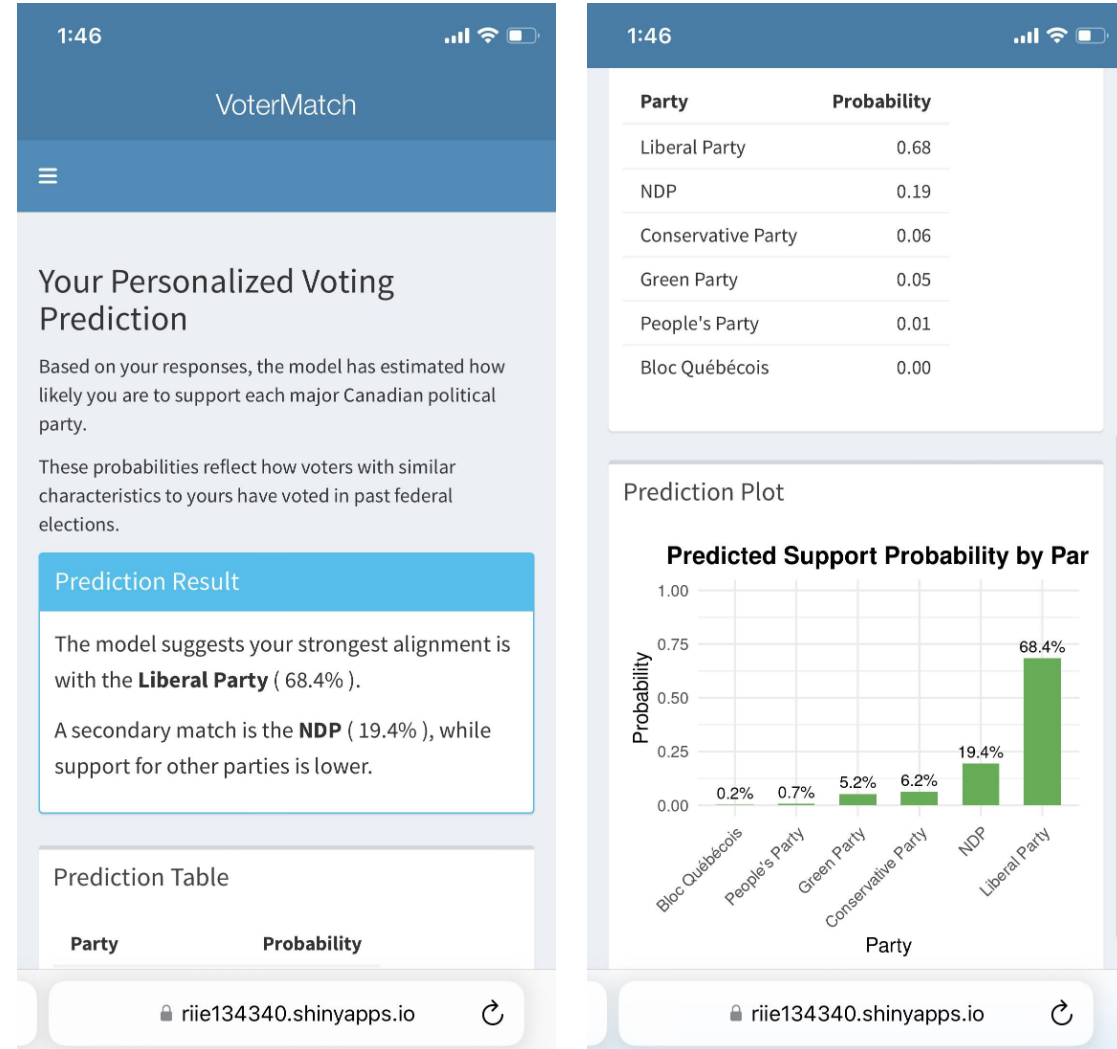
**When there is a conflict between protecting the environment and creating jobs, jobs should come first.**

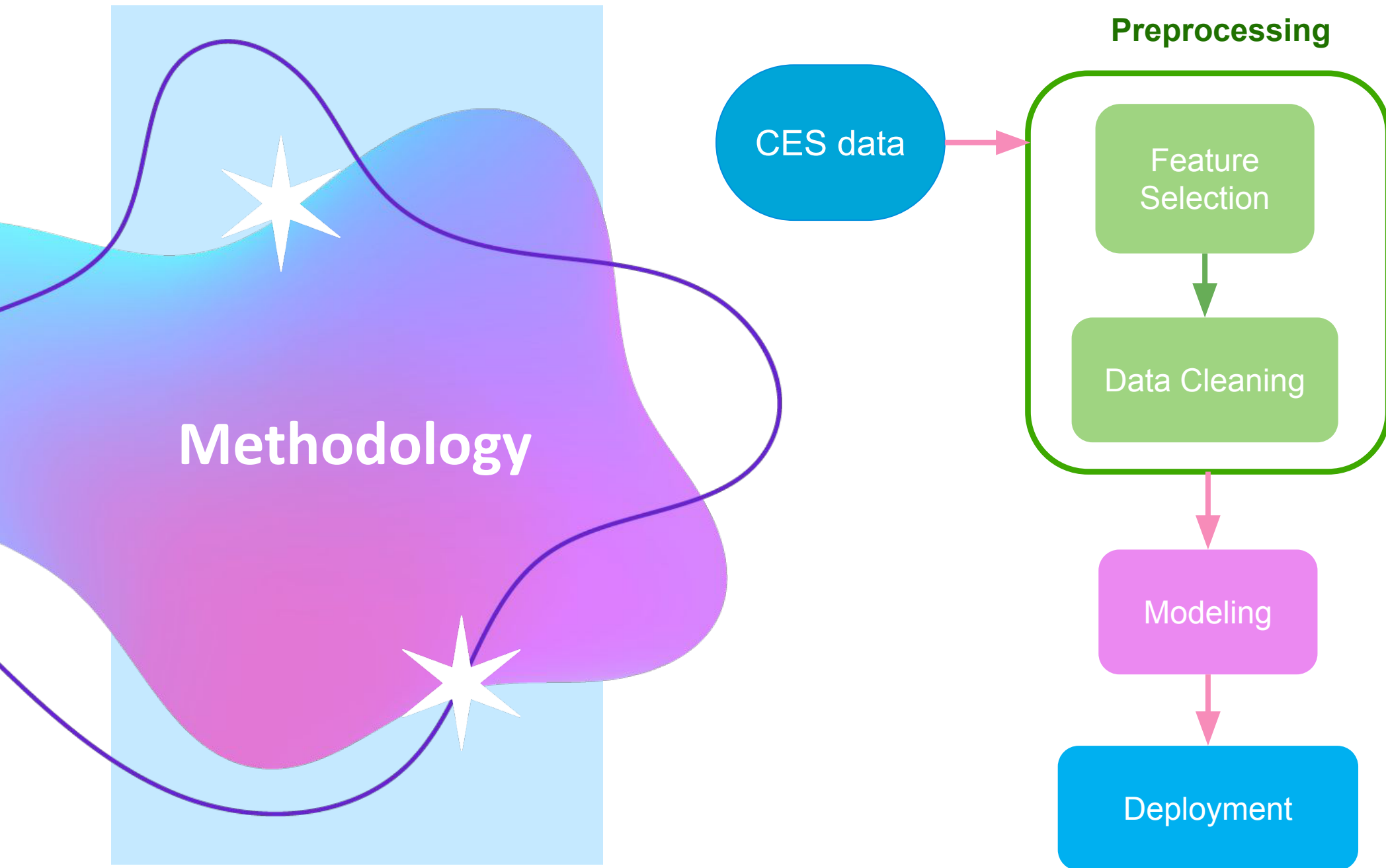
Strongly disagree

Predict

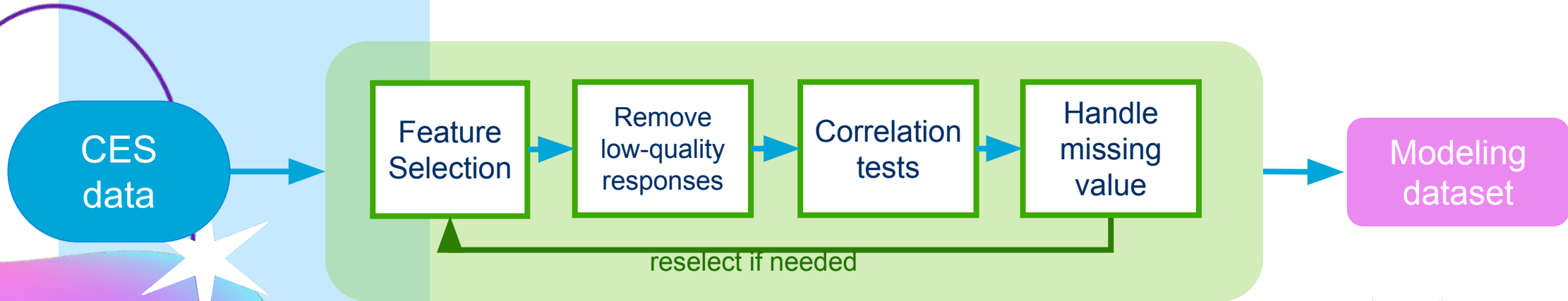
riie134340.shinyapps.io

## Result Page





# Preprocessing



## Feature selection:

- Using correlation tests (Cramér's V & Kruskal–Wallis test)
  - Quickly batch-identify promising variables across many candidates
- Key features: age, province, federal gov't satisfaction, inequality attitude, abortion stance, Canadian-born status, religion importance, marital status, etc.

## Data cleaning:

- Removed low-quality responses (inattentive, duplicates, speeders, spoiled/"don't know" votes)
- Missing Value Handling (~4,000+ rows)



# How the 16 Questions Were Chosen



## Original dataset:

- 1060 variables from CES 2021

## Candidate features:

- personal info,
- policy attitudes
- economic/employment issues,
- a few random picks

Narrowed to 40 candidates after initial review, then applied correlation tests and missing-value filter

# How the 16 Questions Were Chosen

- Were you born in Canada?
- What is your age?
- In which province or territory are you currently reside?
- What is your current marital status?
- Do you think the government does not care much about what people like me think?
- How satisfied are you with the performance of the federal government led by Justin Trudeau?
- Do you think this country has gone too far in promoting equal rights?
- How much do you think the government should do to support racial minorities?
- Do you think abortion should be banned?
- Do you think Canada has gone too far in promoting bilingualism?
- Do you think newer lifestyles are contributing to the breakdown of our society?
- Do you think too many recent immigrants are not trying to integrate into Canadian society?
- Do you think immigrants take jobs away from other Canadians?
- Do you think income inequality is a serious problem in Canada?
- Do you think The government should leave it entirely to the private sector to create jobs?
- When there is a conflict between protecting the environment and creating jobs, jobs should come first?



# Modeling

## Model Selection:

- Tried **Logistic Regression & Random Forest** → struggled with multi-party prediction & minority classes

## Why XGBoost:

- Handles multi-class classification, predicts probabilities for all major parties
- Manages missing values natively, no need for manual imputation
- Works well with categorical features
- Robust to class imbalance with class weights

##	Class: Green Party	Class: Liberal Party	Class: NDP
## Sensitivity	0.00000	0.8385	0.37890
## Specificity	1.00000	0.7935	0.92105
## Pos Pred Value	NaN	0.6830	0.53797
## Neg Pred Value	0.97824	0.9025	0.85940
## Prevalence	0.02176	0.3467	0.19524
## Detection Rate	0.00000	0.2907	0.07398

# Modeling

## Weight Adjustment:

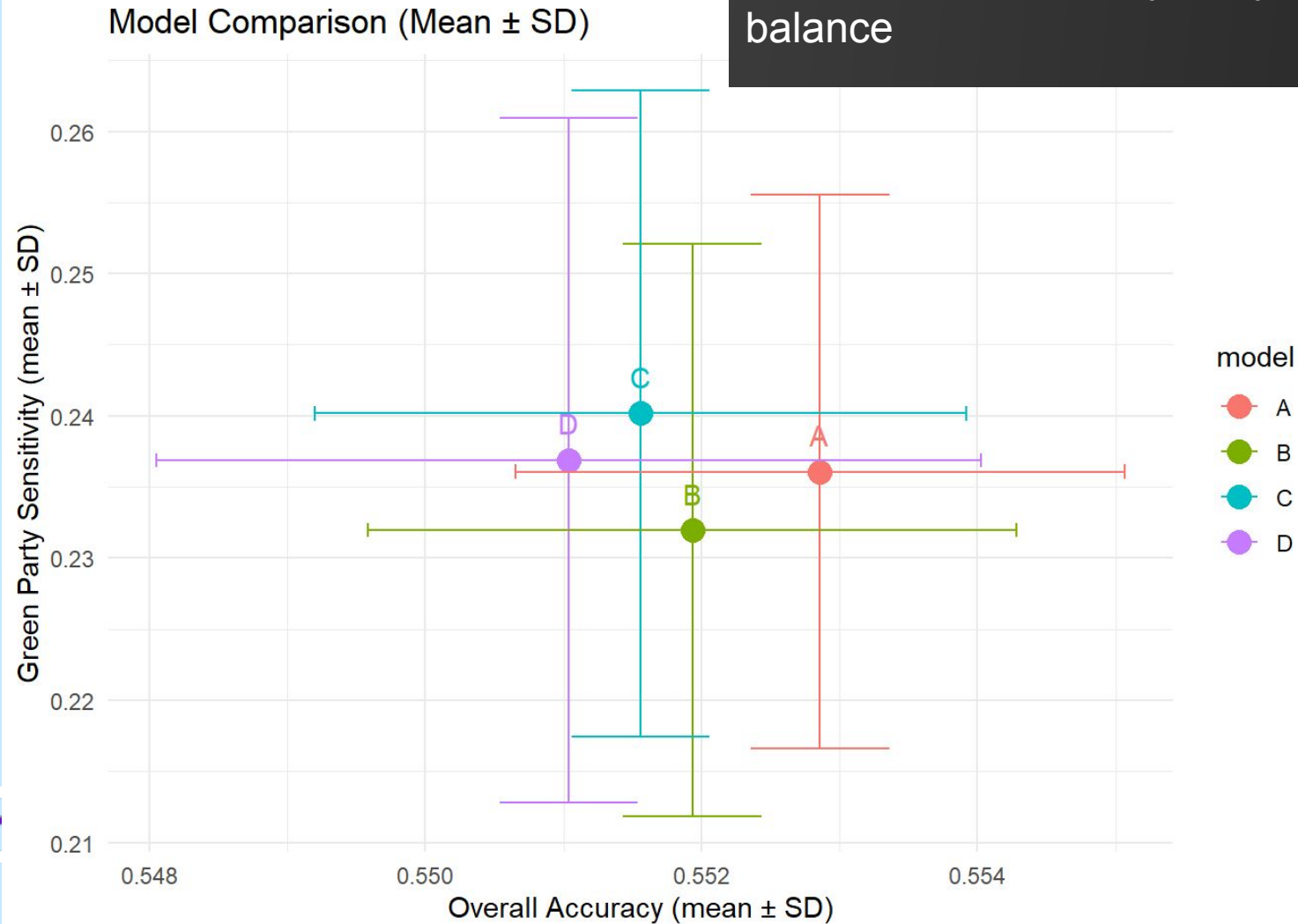
- applied class weights to reduce bias toward larger parties (e.g., Liberal, Conservative) and improve recall for smaller parties

## Tuning

- cross-validation to optimize parameters (e.g., learning rate, depth, rounds) using **accuracy** + **log-loss**

Model	Tuning Metric	Num of rounds	Average Accuracy	Sensitivity (Green)
A	Accuracy	200	0.5528	0.2361
B	Log-loss	200	0.5519	0.2320
C	Accuracy	Best	0.5515	0.2402
D	Log-loss	Best	0.5510	0.2369

Overall accuracy: ~ 55%  
Performance varies by party due to dataset balance



# Model Performance

## Balanced Accuracy:

- Bloc Québécois: 82.7%
- Liberal Party: 76.8%

##	Bloc Québécois	Conservative Party	People's Party
## Sensitivity	0.75781	0.4496	0.64130
## Specificity	0.89677	0.9366	0.90821
## Pos Pred Value	0.51142	0.7314	0.17718
## Neg Pred Value	0.96292	0.8158	0.98797
## Prevalence	0.12480	0.2775	0.02990
## Detection Rate	0.09457	0.1248	0.01917
## Detection Prevalence	0.18492	0.1706	0.10822
## Balanced Accuracy	0.82729	0.6931	0.77476
##	Green Party	Liberal Party	NDP
## Sensitivity	0.245902	0.6569	0.42837
## Specificity	0.932361	0.8798	0.90248
## Pos Pred Value	0.068493	0.7211	0.56309
## Neg Pred Value	0.983905	0.8443	0.84328
## Prevalence	0.019825	0.3211	0.22684
## Detection Rate	0.004875	0.2109	0.09717
## Detection Prevalence	0.071173	0.2925	0.17257
## Balanced Accuracy	0.589131	0.7684	0.66542

# Interactive Tool Demo



SCAN ME

- <https://riie134340.shinyapps.io/shiny/>



# Thank you

Questions?

- [ran@globalfoundation.ca](mailto:ran@globalfoundation.ca)
- [www.linkedin.com/in/ran-shi-a6525b2b5](https://www.linkedin.com/in/ran-shi-a6525b2b5)