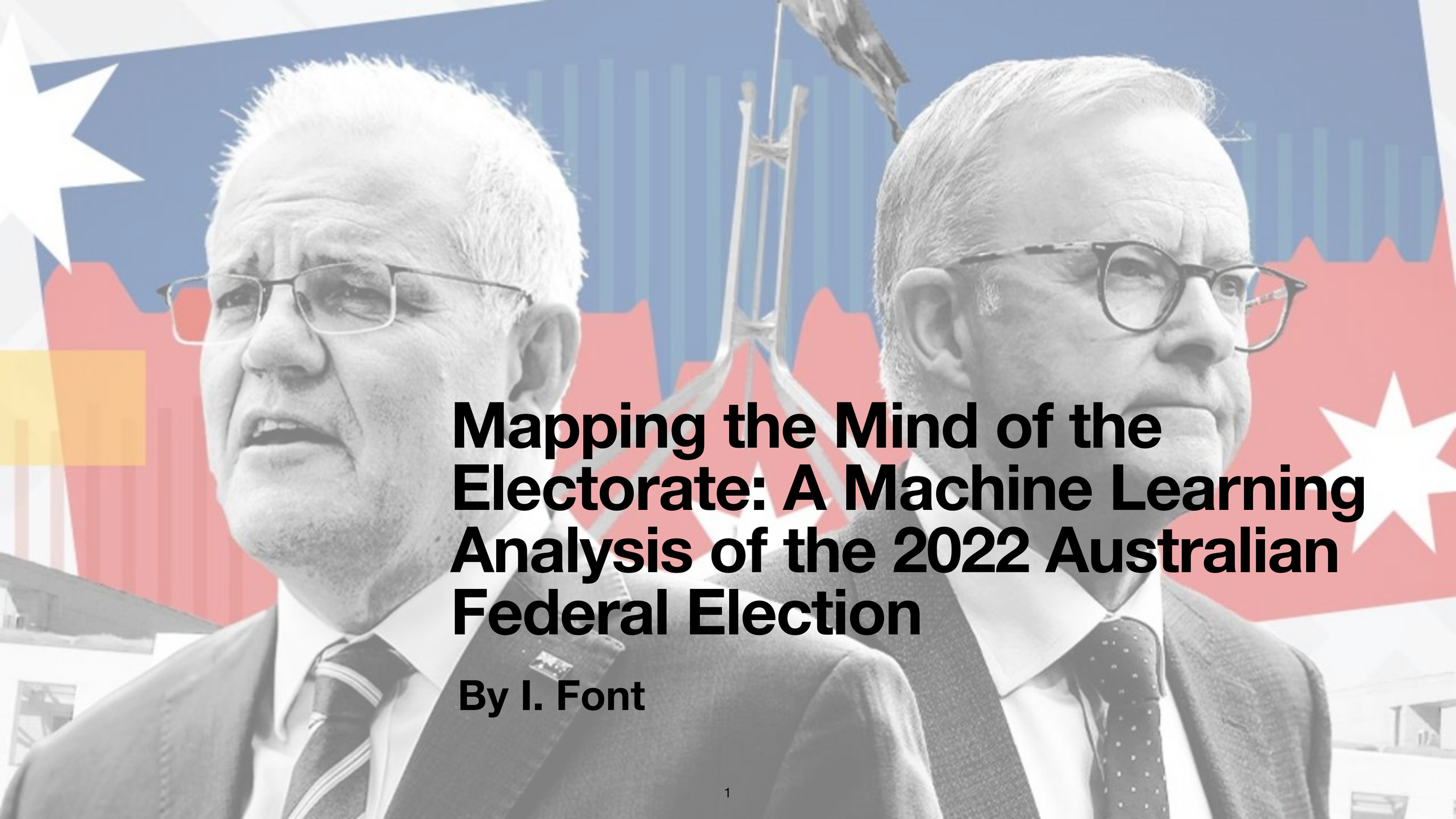
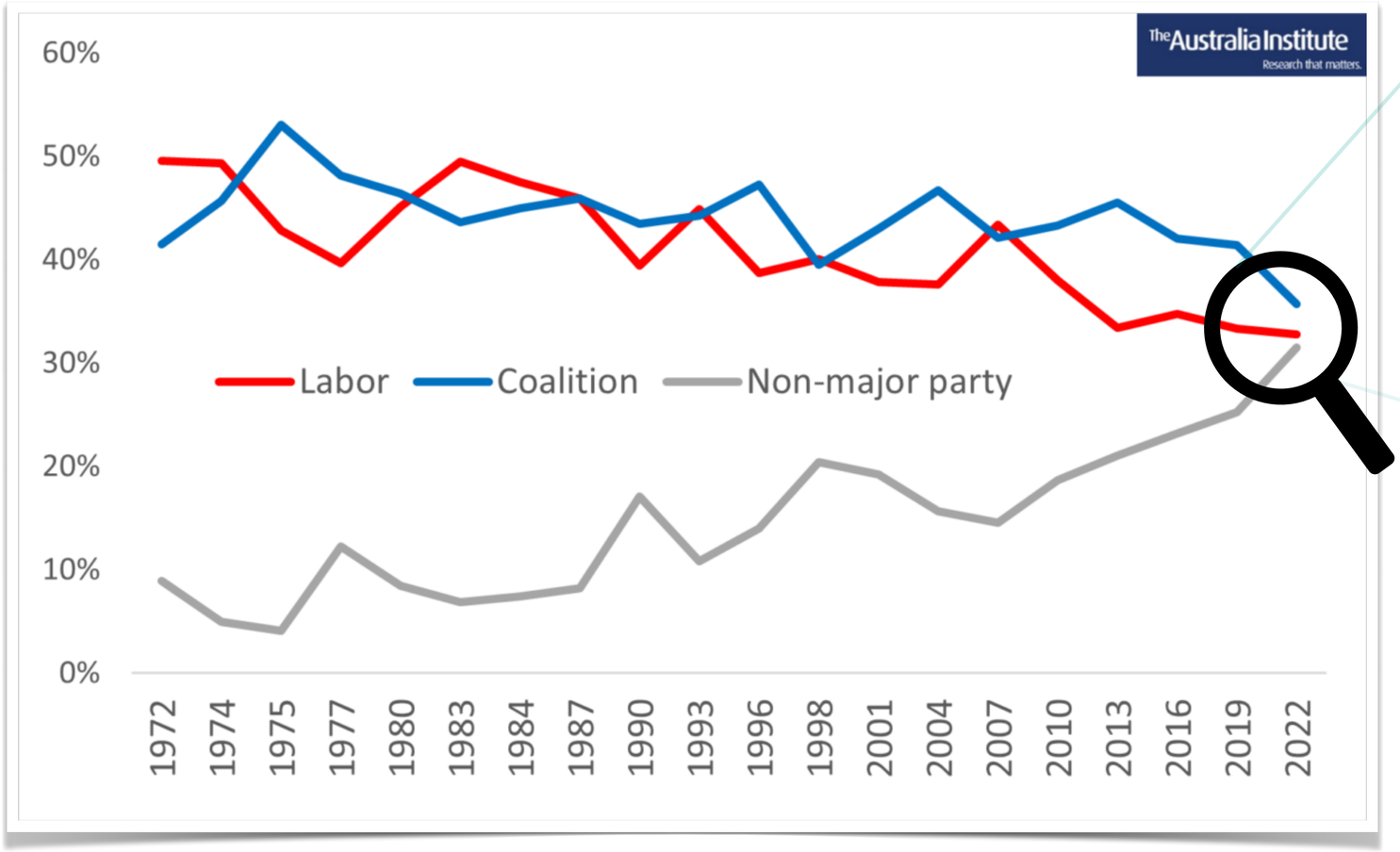


# Mapping the Mind of the Electorate: A Machine Learning Analysis of the 2022 Australian Federal Election

By I. Font

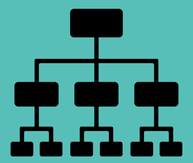
# Why are so many voters leaving the major parties?




**Who are these voters and what do they want?**

*demographics* *values* *priorities*

**Can voter support be predicted with survey data?**



*machine learning*

# Who needs to understand these voters?

Campaign teams

Pollsters

Policy makers

Voter advocacy groups

# The dataset

AUSTRALIAN NATIONAL UNIVERSITY

AES AUSTRALIAN ELECTION STUDY

Survey based rich data

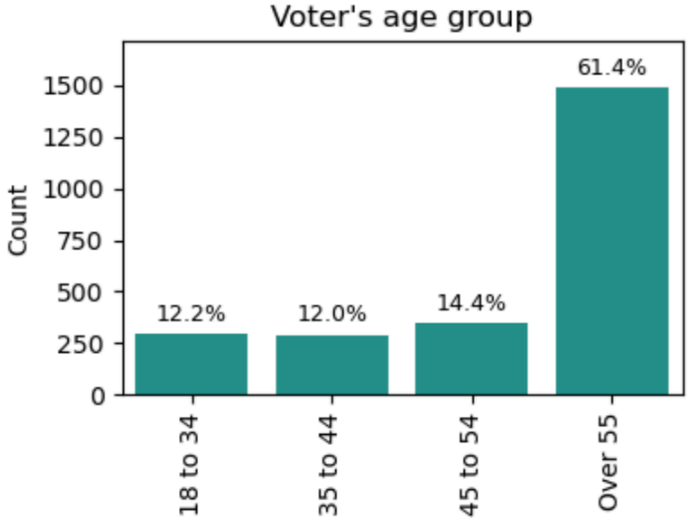
Categorical and ordinal data

2508 samples

350+ features

# What I found

## Age bias



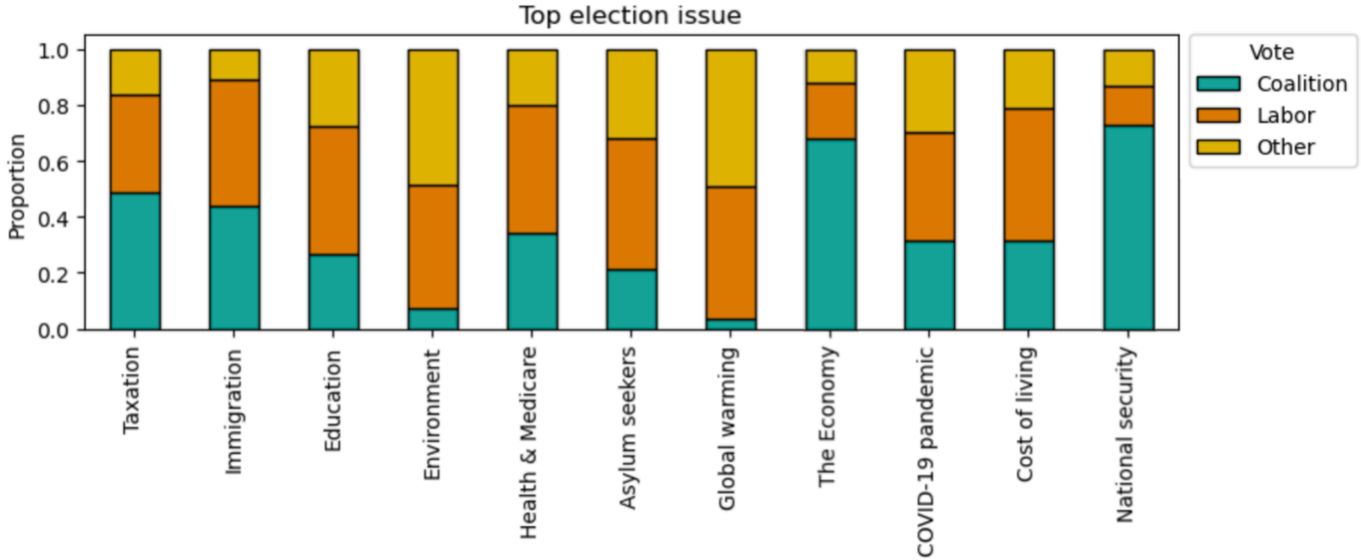

**84 %**
**politically engaged**

Most respondents follow politics closely.

## Party id ≠ votes

More identify with Coalition (37.5% vs 33.7%)
Labor got more votes..! (38.9% vs 35.3%)

## Different priorities



# My approach

# Clustering voters

## Fuzzy c-means

| Demographics | Political attitudes | Spending priorities | Top issues |
|---|---|---|---|

### Progressive voters

health
Concerned **Younger**
skeptical of urban **Climate**
**parties** FEMALE
**Social** SERVICES

### Conservative voters

Economy
**MALE** **police**
Older
more trusting
regional government
Prioritise **DEFENCE**

# From understanding voters to predicting behaviour

**Clustering revealed who voters are…**

**But can we predict how they'll vote**

## Logistic Regression

Most interpretable

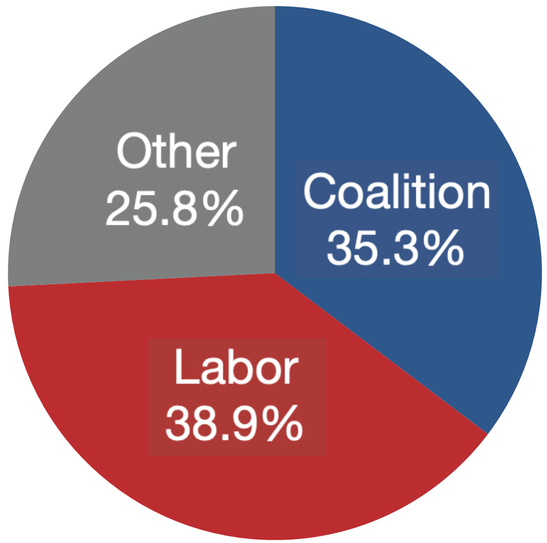
## Random Forest

Robust to noise, finds non-linear splits

## XGBoost

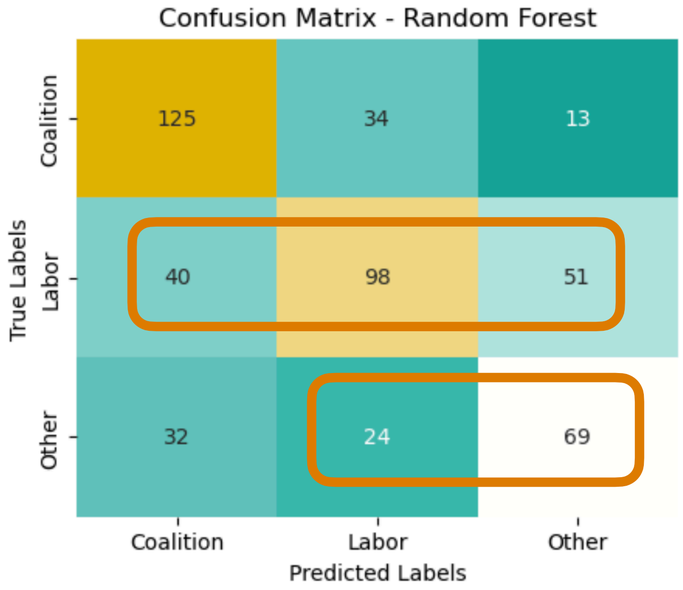
Best accuracy, best separation

Target:
first_preference_vote

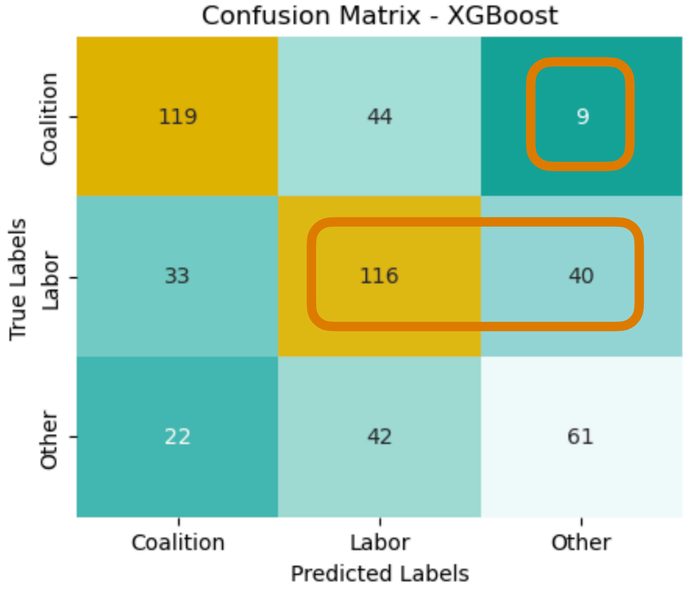


# How well did the models predict voter choice?

LR accuracy = 0.58



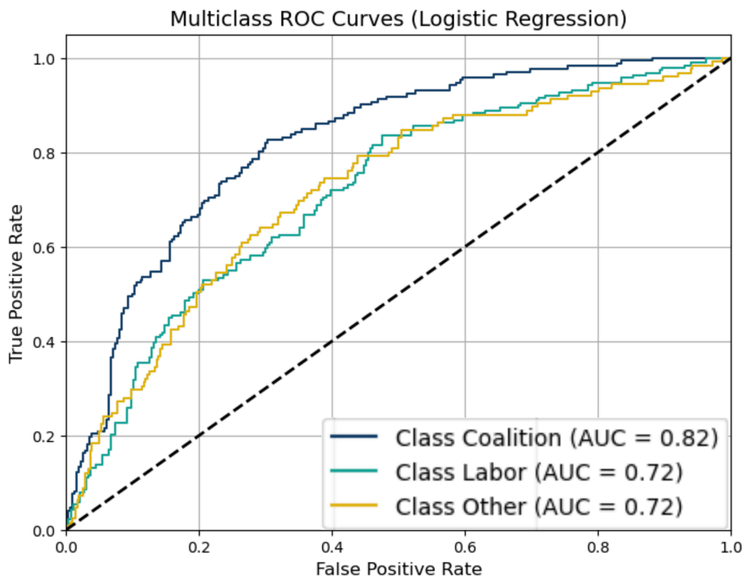

RF accuracy = 0.60

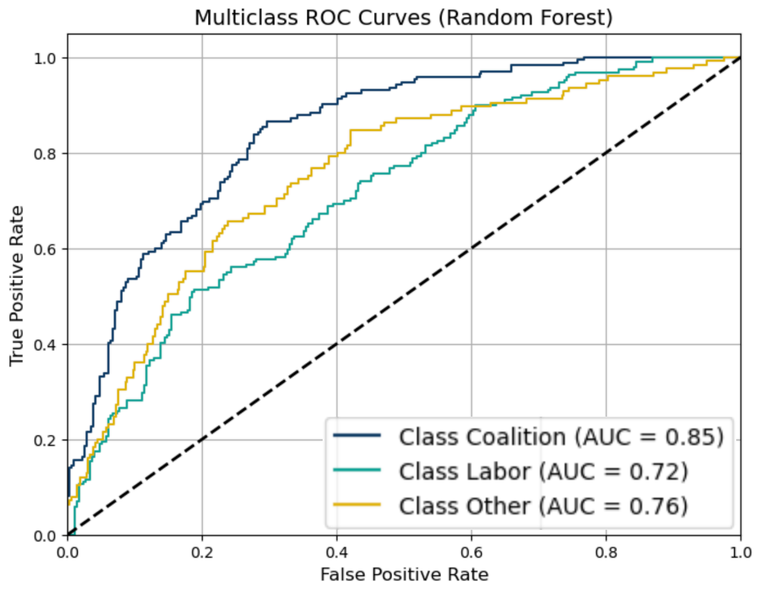


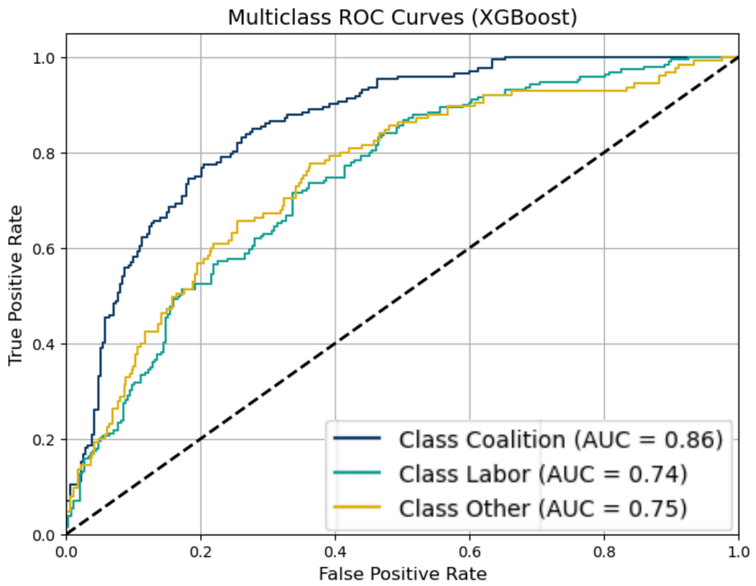

XGB accuracy = 0.61

**XGBoost** and **Random Forest** outperformed Logistic Regression in identifying 'Other' voters — but none achieved strong predictive power.

# How confident were the models in their predictions?




**XGBoost** showed the best class separability, especially for Coalition voters, but still struggled with nuanced classes like ‘Other’.

# What the data really tells us

Data shows how voter segments think

but not always how they vote.

# Where this can go next

Compare AES waves over time

Use more features

Balance voter ages

**This is just one election.**
**What could we learn from five?**

Questions?