

# Predicting Political Behavior Using Digital Trace Data

Ruben Bach, Christoph Kern, Ashley Amaya, Florian Keusch, Frauke Kreuter, Jan Hecht, and Jonathan Heinemann

## 1) Motivation

Previous studies used digital trace data to

- predict socio-demographic characteristics
- forecast election results

Test our ability to predict *individual* voting behavior in 2017 German election using digital trace data, survey data, and a combination of the two

## 3) Method

Series of Gradient Boosting Machine models (Friedman 2001, Friedman et al. 2000) as implemented in XGBoost (Chen and Guestrin 2016)

Hyperparameters tuned based on grid search

- 10-fold CV
- 75% training data
- Maximizing ROC-AUC

Final models built on full training data and evaluated on 25% test data using ROC curves, ROC AUC

## 2) Data

Records of individual browsing behavior and app usage, provided by a German non-probability panel

Three waves of survey data

- Voting intentions (prior to the election) and actual voting decision (after the election)
- Consumption of political information
- several socio-demographic variables

*Outcomes*

1. Undecided (party) prior to election
2. Voted in election
3. Voted for AfD (right-wing populist party)

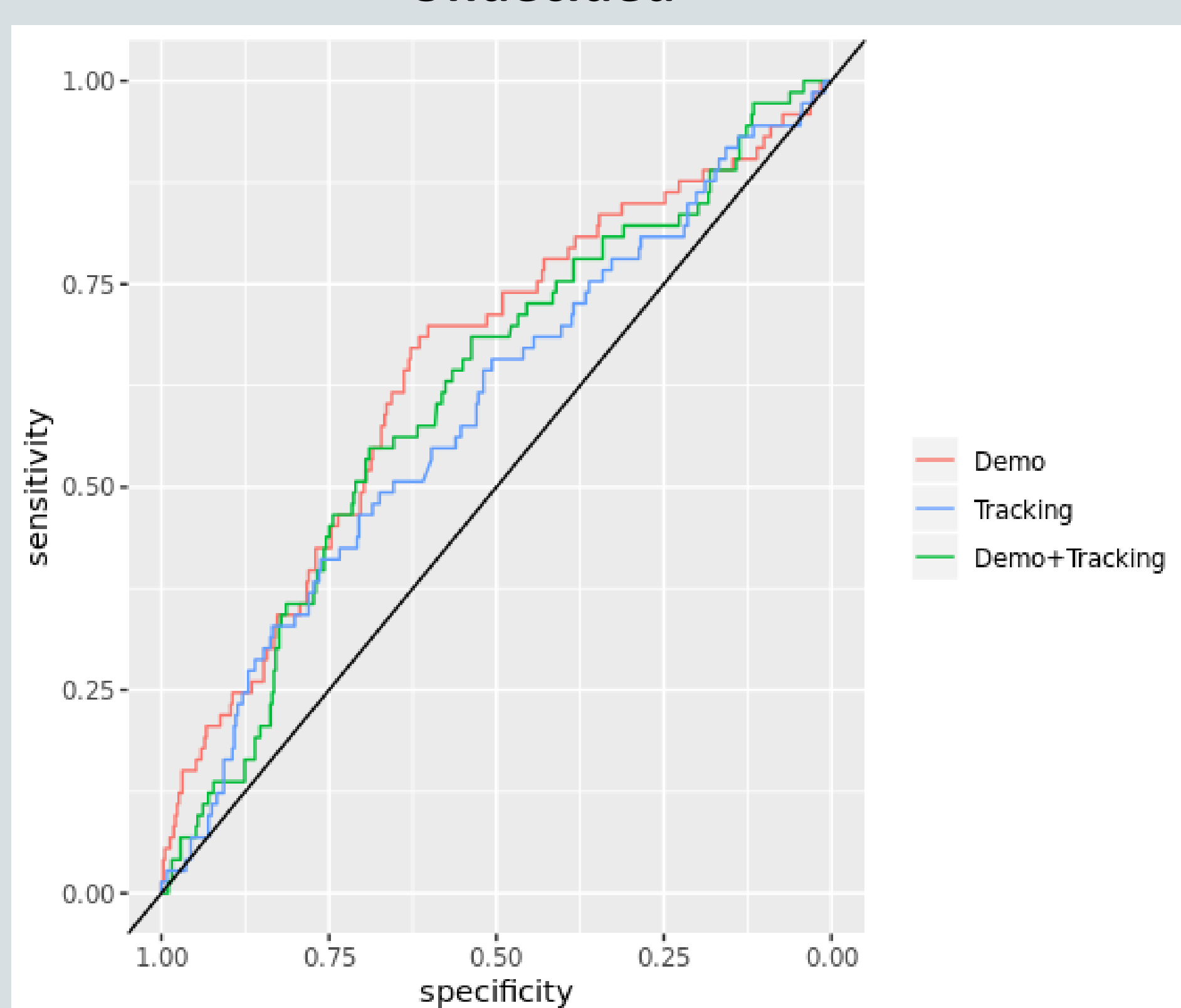
*Predictors (about 11,000)*

- Information on browsing behavior and app usage (e.g., duration and devices used)
- Usage of apps and categories of apps
- Consumption of mainstream news media and \*alternative\* media (e.g., right-wing populist online blogs)
- Domain visited
- Socio-demographic predictors (e.g., age, gender, income, education) from survey data

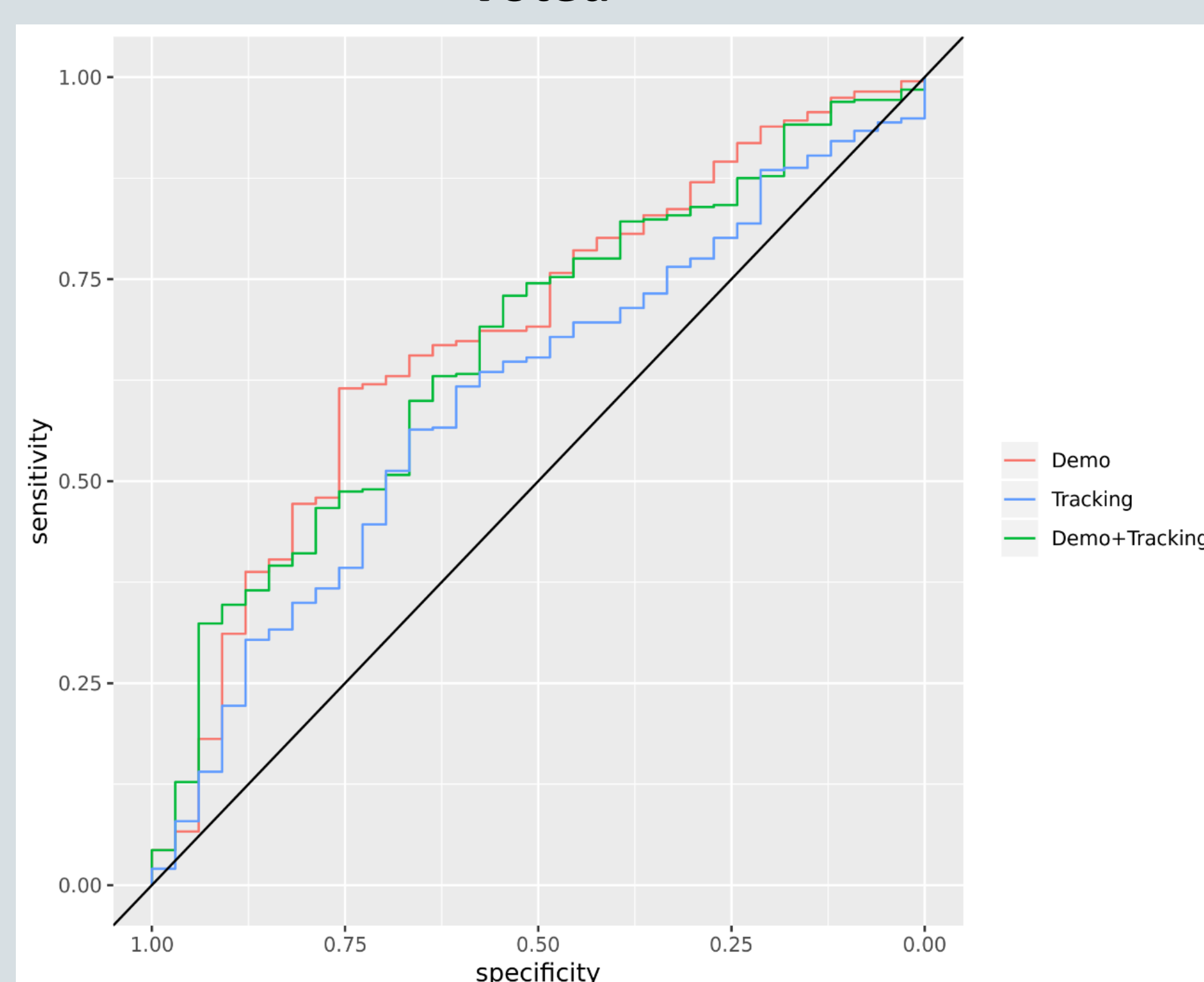
	Yes	No	Total
Undecided	344 (19.1%)	1,455 (80.9%)	1,799
Voted	1,590 (92.7%)	126 (7.3%)	1,716
Voted for AfD	224 (14.6%)	1,308 (85.4%)	1,532

## 4) Results

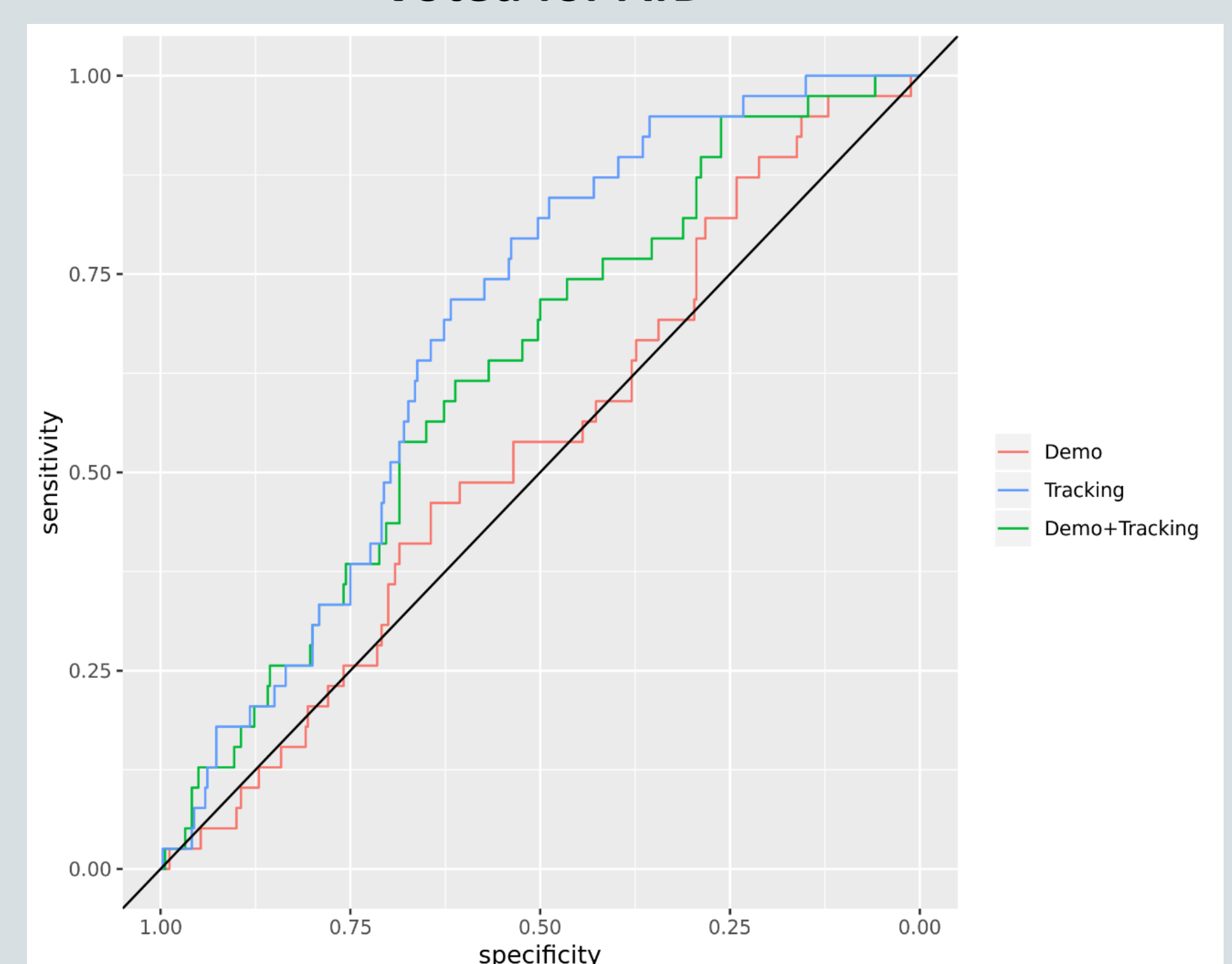
Undecided



Voted



Voted for AfD



- Using digital trace data instead of socio-demographic information to predict political behavior and attitudes results in comparable predictive performance
- Combining the two sources of data does not increase predictive performance
- Using records of online behavior seems especially promising regarding party preferences (voted for AfD)

## 5) Conclusion

Although better than random guessing, model performance does not allow precise prediction of political behaviors or attitudes (e.g., to replace surveys), but may be useful for targeting political campaigns

Future steps

- Extract search terms from online searches and incorporate information in prediction models
- Extract more information on social media activity from URLs (e.g., group membership)

### CONTACT

Ruben Bach

(r.bach@uni-mannheim.de)

SFB 884, Political Economy of  
Reforms

University of Mannheim