Can online behavior predict political attitudes?

Paulina Pankowska[1],Davide Morselli[2], and Ruben Bach[3]

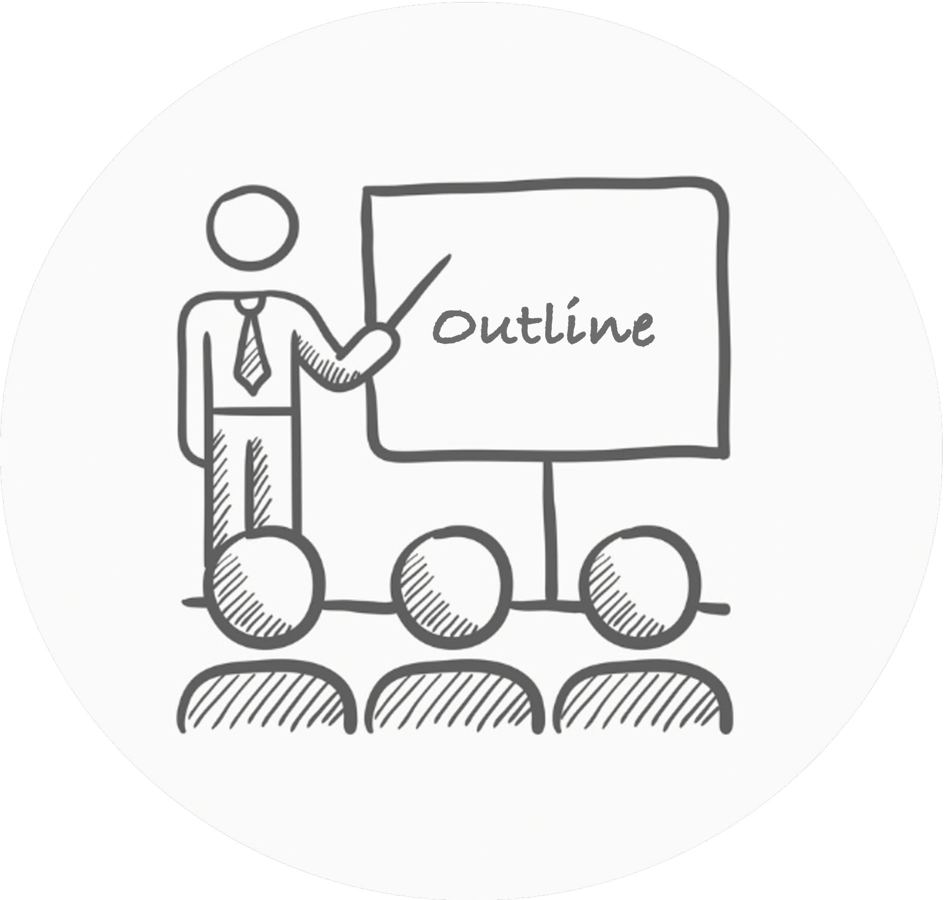
[1] Sociology Department, Utrecht University, The Netherlands
[2] LIVES Centre, University of Lausanne, Switzerland
[3] MZES Data and Methods Unit, University of Mannheim, Germany

17th Conference “Social Monitoring and Reporting in Europe”
9-11 October 2023
Villa Vigoni

Utrecht University

LIVES SWISS CENTRE OF EXPERTISE IN LIFE COURSE RESEARCH

UNIVERSITY OF MANNHEIM

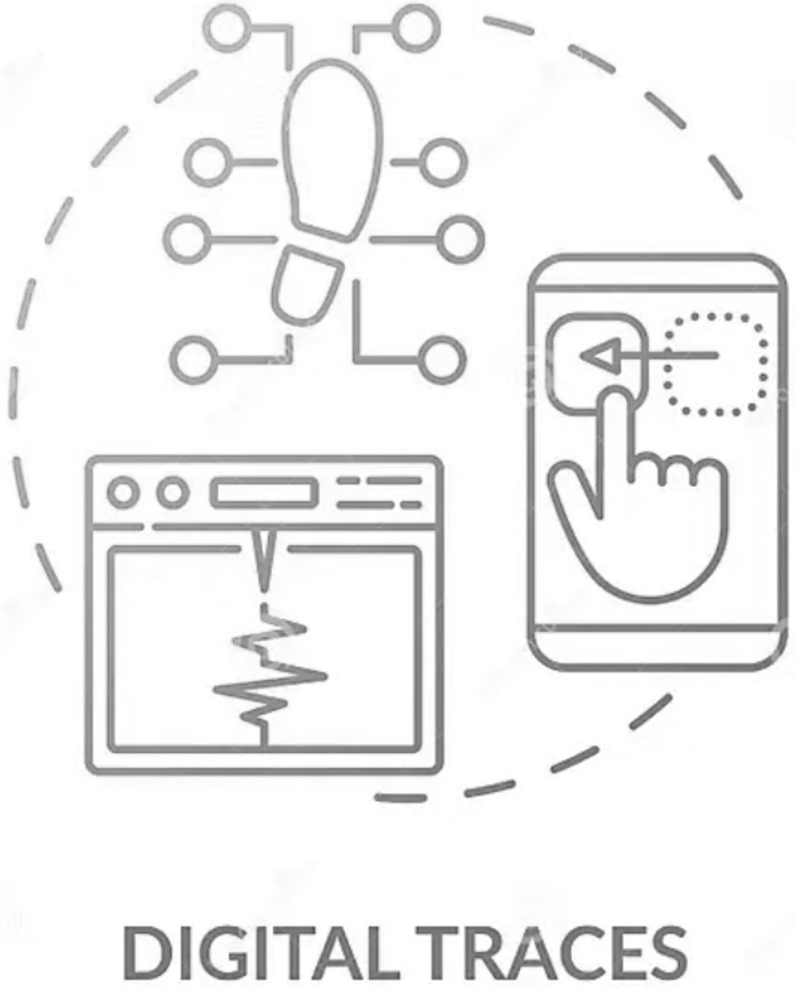


Background

- *Methodological perspective:* increased use of digital trace data in social sciences
- *Substantive perspective:* relationship between online browsing and political orientation

Our research

- Predicting self-reported political orientation using browsing data



- Recent increased interest in the social sciences
- Considered an attractive alternative to surveys
  - Superior quality
  - Allow to overcome limitations
- Have limitations that are often ignored
  - Representativeness
  - Demographics, attitudes, beliefs have to be inferred

- Increased interest in predicting individual level characteristics
  - Focus on social media or media consumption
- Predicting demographics
  - Overall high accuracy (e.g, gender or age)
- Predicting attitudes, opinions, beliefs…
  - More mixed picture

Relationship between what people read online and their political orientation

- Confirmation bias
- Echo chambers
- Filter bubbles

Can we predict political orientation using information on websites visited and apps used?

- Outcome: political orientation - *from survey*
- Predictors: categories of websites browsed, and apps used - *from tracking apps*
- Machine learning approach

Non-probability online panel in Germany

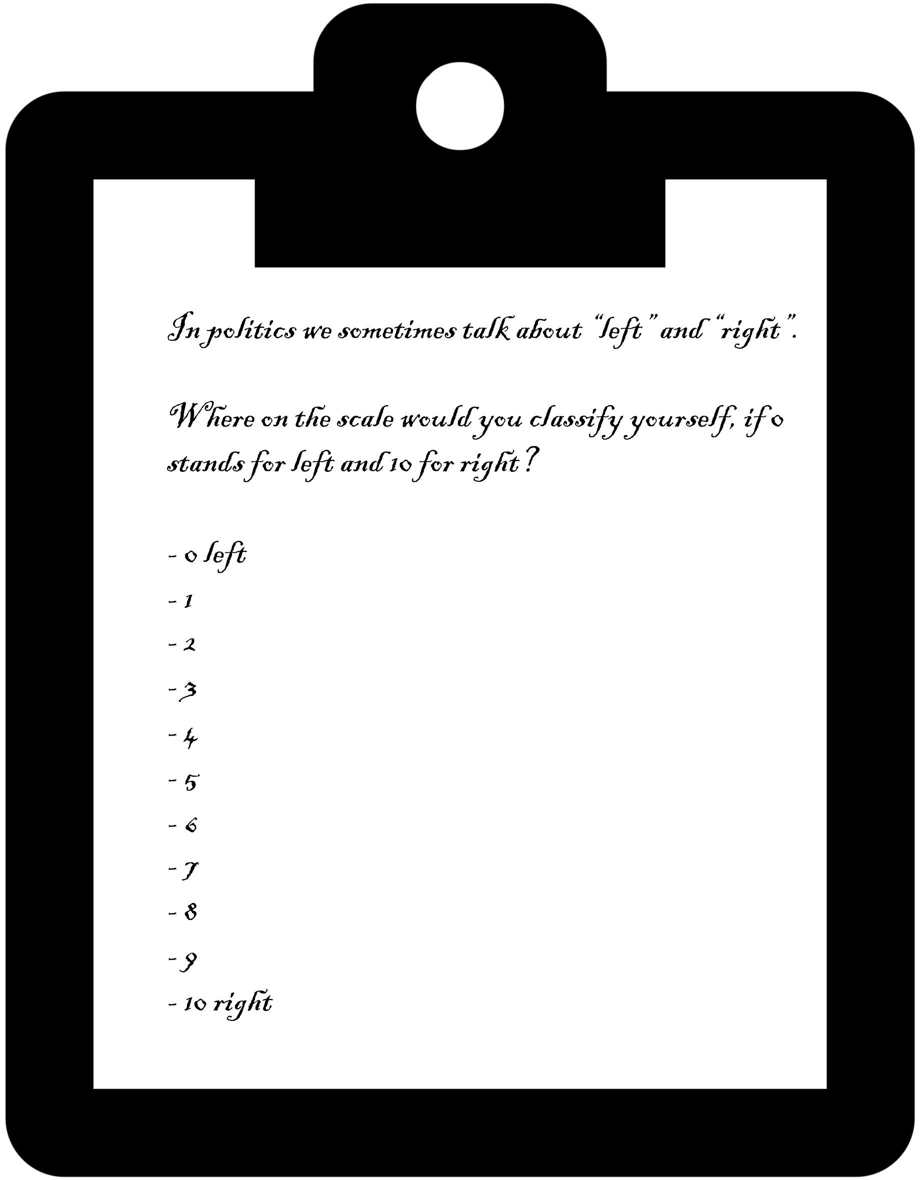
- N=2,100
- Survey with three waves around September 2021
- Tracking app installed on PC and/or mobile device
- > 600 respondents had both PC and mobile tracked

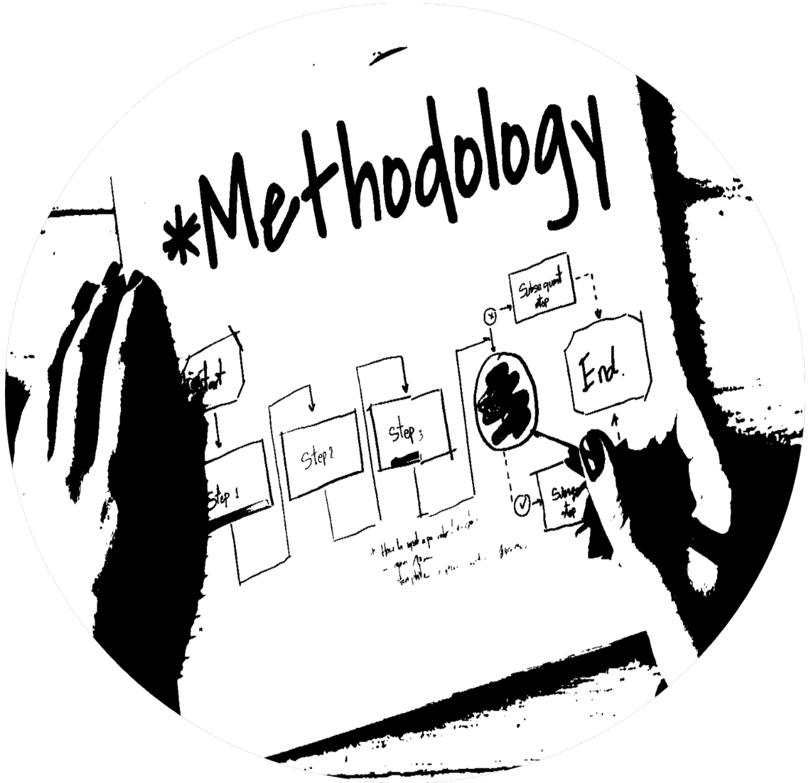
Survey

- 11-point Likert scale asking about political orientation

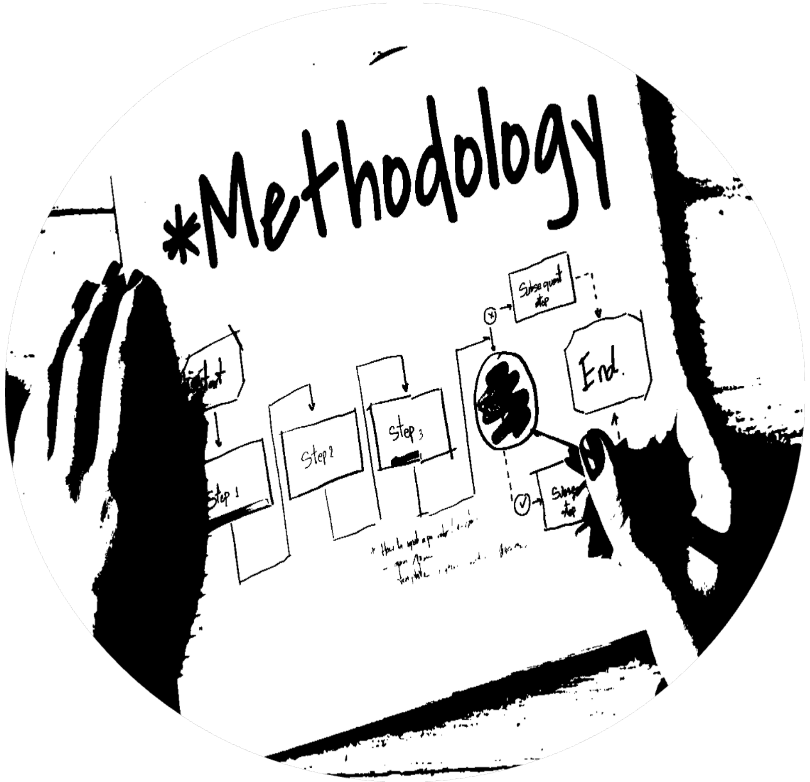
Tracking app

- Mobile, PC, or both
- Total period up to 6 months (Jul – Dec ‘21)
- Information used on domains browsed on PC/mobile and apps used on mobile





- Classification of web domains and apps using ChatGPT
- Looking at usage from PC & mobile separately
- Two operalizations of usage: count and duration



Machine learning approach

- Random forest
- 70% of sample used to train model
- 30% of sample use to validate predictions

Baseline model (age, gender, immigration status, edu) vs. tracking | baseline & tracking

Baseline model

$MSE_{SD} = 1.91$

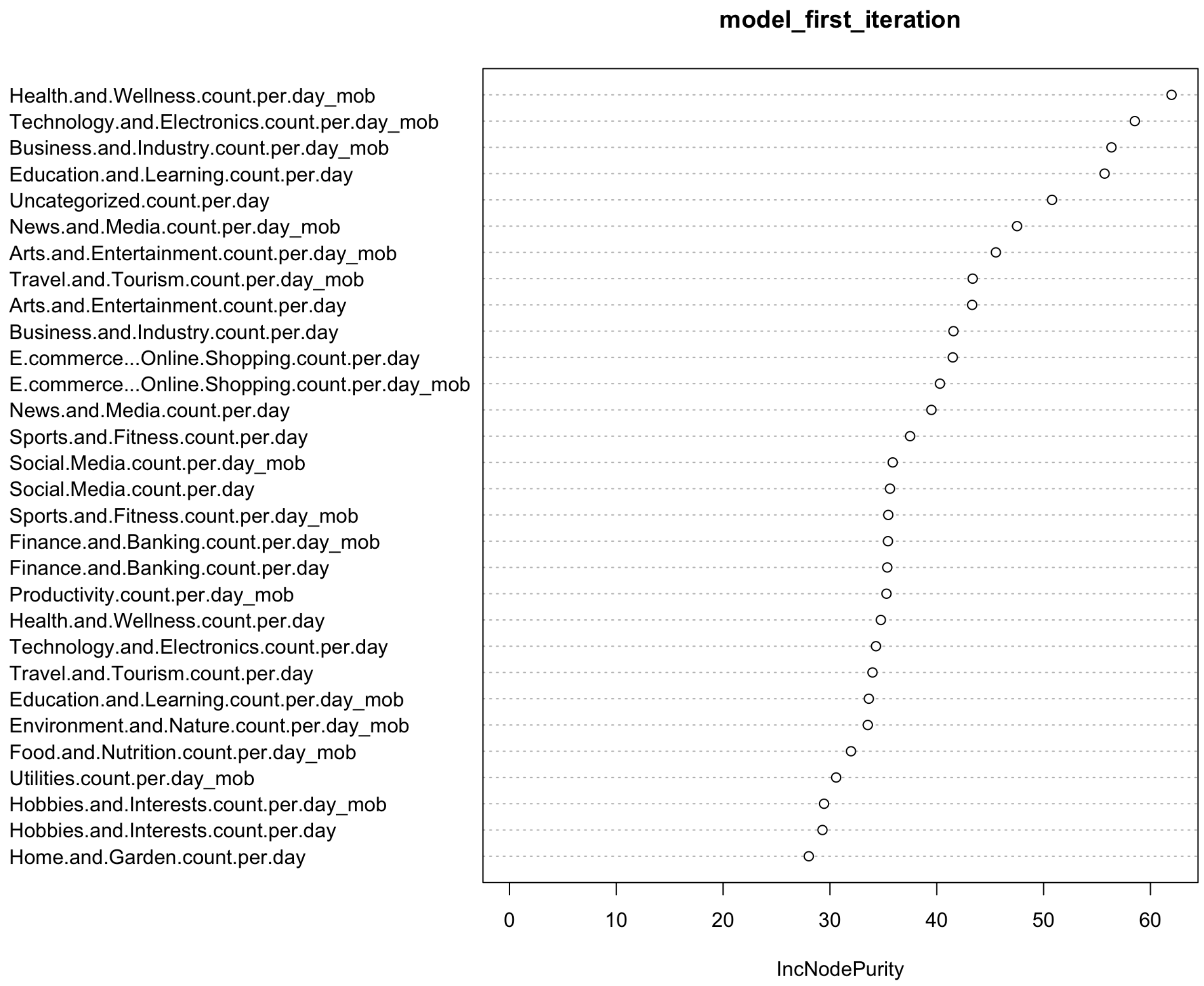
Count/duration per day (mobile & PC)

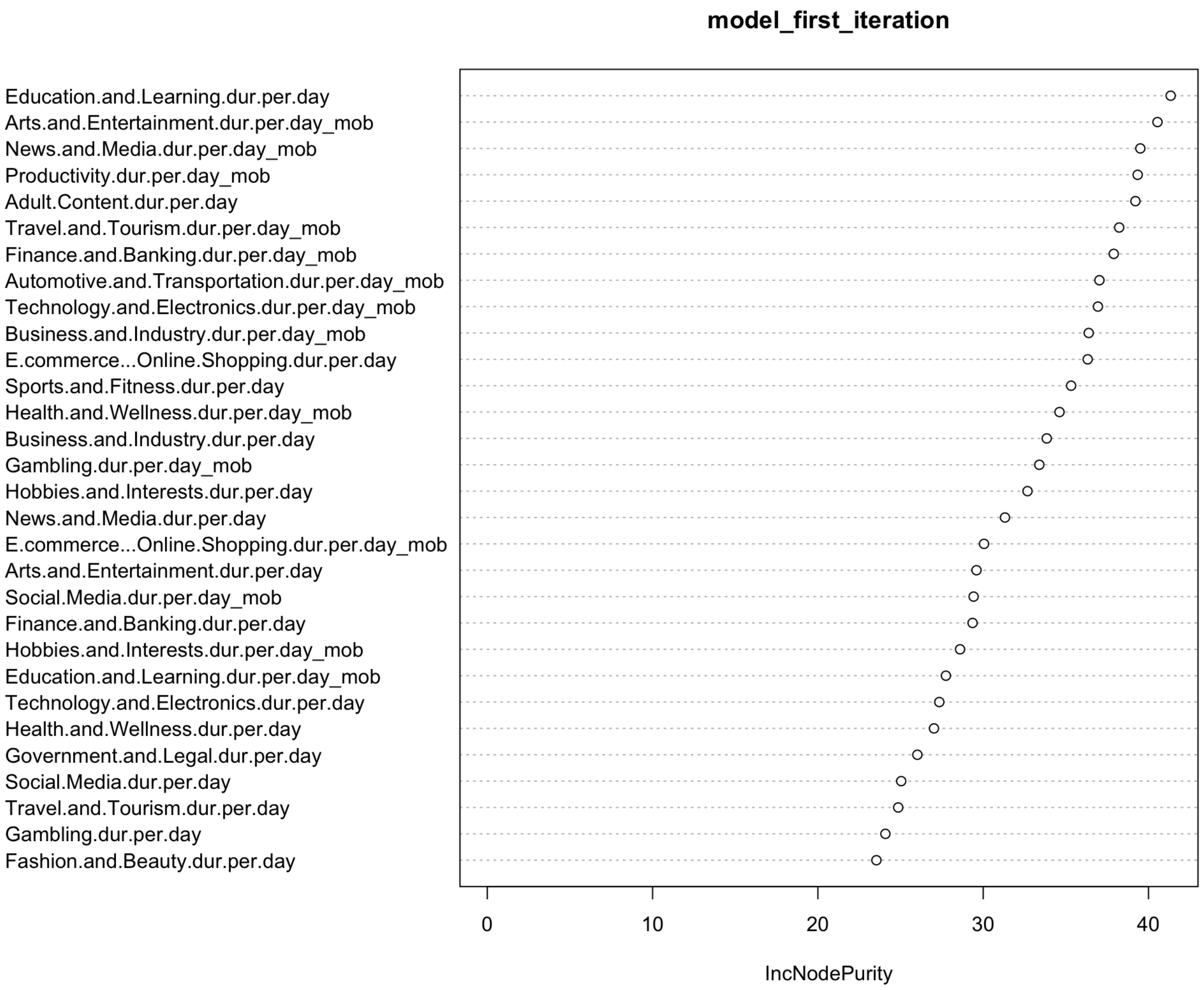
$MSE_{SD} = 1.89$

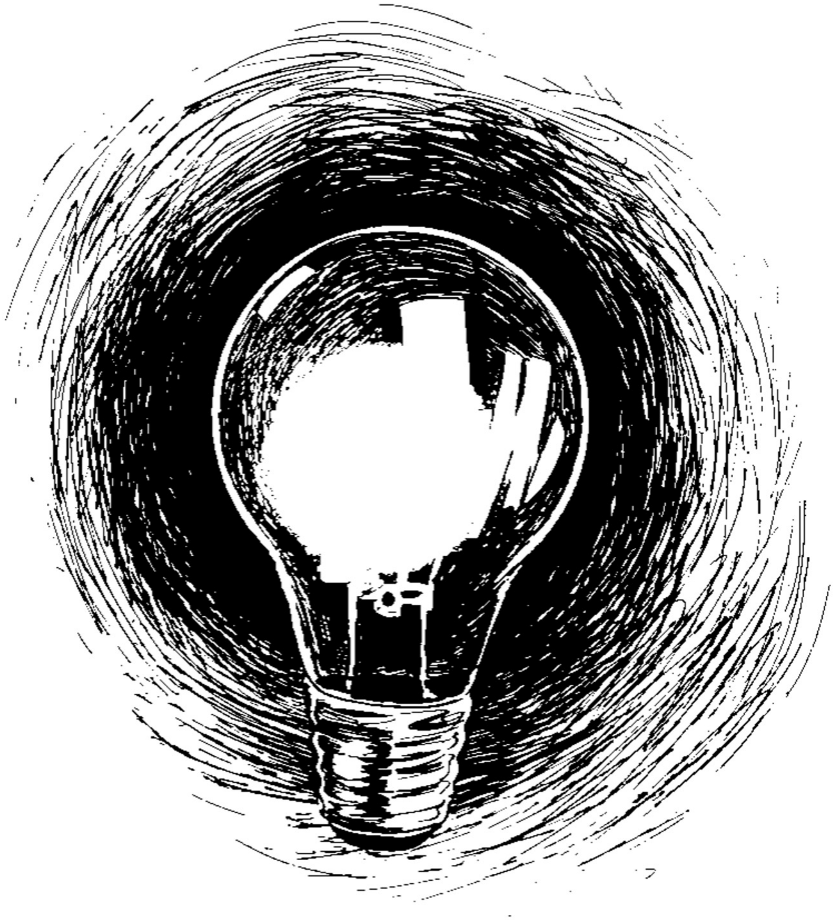
Baseline & duration/count (mobile & PC)

$MSE_{SD} = 1.91$

Differences statistically **insignificant**


model_first_iteration
Health.and.Wellness.count.per.day_mob
Technology.and.Electronics.count.per.day_mob
Business.and.Industry.count.per.day_mob
Education.and.Learning.count.per.day
Uncategorized.count.per.day
News.and.Media.count.per.day_mob
Arts.and.Entertainment.count.per.day_mob
Travel.and.Tourism.count.per.day_mob
Arts.and.Entertainment.count.per.day
Business.and.Industry.count.per.day
E.commerce...Online.Shopping.count.per.day
E.commerce...Online.Shopping.count.per.day_mob
News.and.Media.count.per.day
Sports.and.Fitness.count.per.day
Social.Media.count.per.day_mob
Social.Media.count.per.day
Sports.and.Fitness.count.per.day_mob
Finance.and.Banking.count.per.day_mob
Finance.and.Banking.count.per.day
Productivity.count.per.day_mob
Health.and.Wellness.count.per.day
Technology.and.Electronics.count.per.day
Travel.and.Tourism.count.per.day
Education.and.Learning.count.per.day_mob
Environment.and.Nature.count.per.day_mob
Food.and.Nutrition.count.per.day_mob
Utilities.count.per.day_mob
Hobbies.and.Interests.count.per.day_mob
Hobbies.and.Interests.count.per.day
Home.and.Garden.count.per.day
0
10
20
30
40
50
60
IncNodePurity


model_first_iteration
Education.and.Learning.dur.per.day
Arts.and.Entertainment.dur.per.day_mob
News.and.Media.dur.per.day_mob
Productivity.dur.per.day_mob
Adult.Content.dur.per.day
Travel.and.Tourism.dur.per.day_mob
Finance.and.Banking.dur.per.day_mob
Automotive.and.Transportation.dur.per.day_mob
Technology.and.Electronics.dur.per.day_mob
Business.and.Industry.dur.per.day_mob
E.commerce...Online.Shopping.dur.per.day
Sports.and.Fitness.dur.per.day
Health.and.Wellness.dur.per.day_mob
Business.and.Industry.dur.per.day
Gambling.dur.per.day_mob
Hobbies.and.Interests.dur.per.day
News.and.Media.dur.per.day
E.commerce...Online.Shopping.dur.per.day_mob
Arts.and.Entertainment.dur.per.day
Social.Media.dur.per.day_mob
Finance.and.Banking.dur.per.day
Hobbies.and.Interests.dur.per.day_mob
Education.and.Learning.dur.per.day_mob
Technology.and.Electronics.dur.per.day
Health.and.Wellness.dur.per.day
Government.and.Legal.dur.per.day
Social.Media.dur.per.day
Travel.and.Tourism.dur.per.day
Gambling.dur.per.day
Fashion.and.Beauty.dur.per.day
0
10
20
30
40
IncNodePurity



Conclusions

- Adding tracking data does not improve political orientation prediction
- Prediction accuracy using tracking data only similar to basic demographics only
- Better classification of domains and apps and more diversity in outcome might improve prediction accuracy

- Funding for this study come from the *Baden-Württemberg Foundation* through the grant “Filter Bubbles, Alternative News and Political Polarization” (PI: R. Bach) and the *German Science Foundation (DFG)* through the grant “Evaluating Data Sources for Research into Political Reforms: (Non)probability Online Surveys and Big Data” (PIs: A. Blom, F. Keusch, & F. Kreuter).

- We thank Ella Häcker and Leonard Bek for research assistance

- Illustrations taken from freepik.com and 123rf.com

Thank you!

Dr Paulina Pankowska | Assistant Professor | Utrecht University | Sociology
p.k.pankowska@uu.nl | https://www.uu.nl/medewerkers/PKPankowska