

# Using Supervised Classification for Categorizing Answers to an Open-Ended Question on Survey Motivation

**Carolina Haensch, Bernd Weiß**  
Katja Bitz, Patricia Steins, Priscilla Chyrva

12.05.2022





LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# 1. Background



# The GESIS Panel

- Mixed-mode (online/mail) panel study
- 5000 panelists in Germany
- Running since 2014
- Up to six waves per year
  
- Annual open-ended question about panelists' motivation to participate
  - "For what reasons do you participate in the surveys of the GESIS Panel?"
  - More than 25.000 collected answers

The recruiter was nice

It's fun!

I like to give my opinion

It's nice to receive 5 Euros for participating

I want to influence those in power



# Motivation



## Panel management

Survey attitudes are predictors of panel attrition  
(Groves et al. 2000; Schnell 1997)



## Existing scales to capture survey motivation

Might not capture the entire space of motivations



## Open-ended question

Allow panelists to voice their opinion without restrictions  
Time-intensive to analyze

# Research Questions

## 1. Methodological questions

1.1 How can the reasons for participation be grouped?

1.2 Given a coding scheme, how can we classify 25000+ answers?

## 2. Substantive questions

2.1 Which motivations for participation do GESIS panelists show?

2.2 What is the predictive value of the new variable?





## 2. Methodological Questions

Semi-automatic classification of answers



# Categories of Panel Motivation

## Existing classifications for open-ended survey motivation questions:

- Dividing reasons into altruistic, egoistic reasons and reasons related to survey characteristics (Singer 2011)
- Our starting point

## Own classification:

- 21 categories
- Iteratively developed at the GESIS Panel
- Hand-coded 5000 answers (two raters)
  - Cohen's Kappa: 0.91



# Categories of Panel Motivation

- Interest
- Curiosity
- Learning
- Tell opinion
- Influence
- Incentive
- Fun
- Routine
- Dutifulness
- Help science
- Help politicians
- Help society
- Help, not further specified



- Brevity
- Anonymity
- Professionalism
- Recruiter
- Recruitment
- Other survey characteristics
- Importance in general
- No reason
- Non-identifiable/Other





**(2) Aus welchen Gründen nehmen Sie an den Umfragen des GESIS GesellschaftsMonitors teil?**

Bitte nennen Sie die drei wichtigsten Gründe.

Wichtigster Grund:

---

Zweitwichtigster Grund:

---

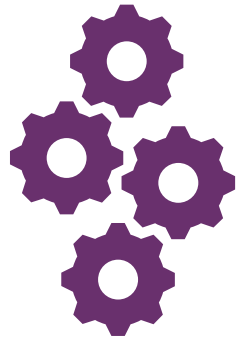
Drittwichtigster Grund:

---

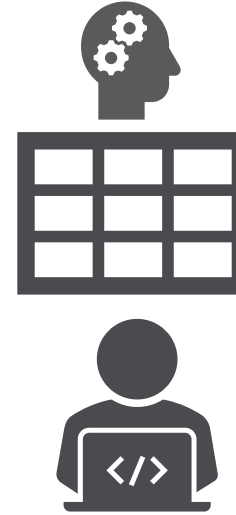
# Supervised Learning



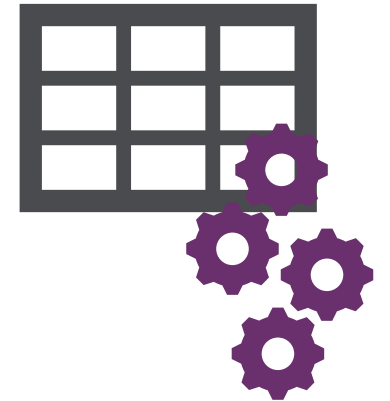
Create data set with  
handcoded  
classification



Train ML  
model



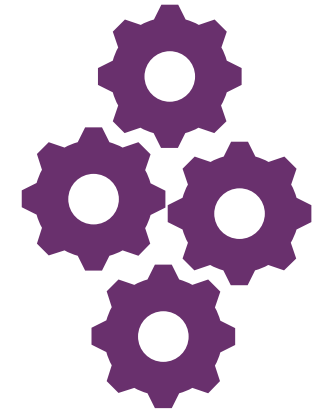
Check  
performance with  
validation set



Classify remaining  
answers through  
the trained model



# Supervised Learning



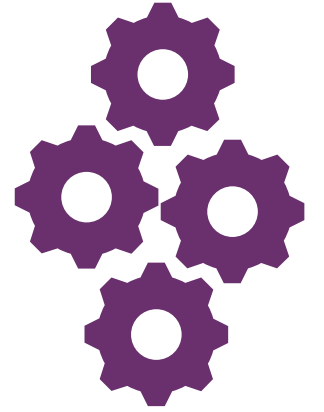
- **LiblineaR** (Helleputte 2017) based on LIBLINEAR (Fan et al. 2008)
- The search for the best model was done via 5-fold cross-validation
  - All model types from LiblineaR
  - Cost parameters 0.001, 1, 10, 1000
- Final specifications
  - L1-regularized L2-loss Support Vector Classification
  - C-parameter: 10



# Supervised Learning

## Evaluation:

- Good performance in general, but a few categories with bad micro measures
  - Relatively fluid and small categories are problematic
  - Entries with typos were hard to predict
    - Added spell checker hunspell (Ooms 2020)
    - Included Word/Sentence Embeddings (FastText/GloVe)
- Additional improvements in performance were small





LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# 3. Analytical Value



# Descriptive Results

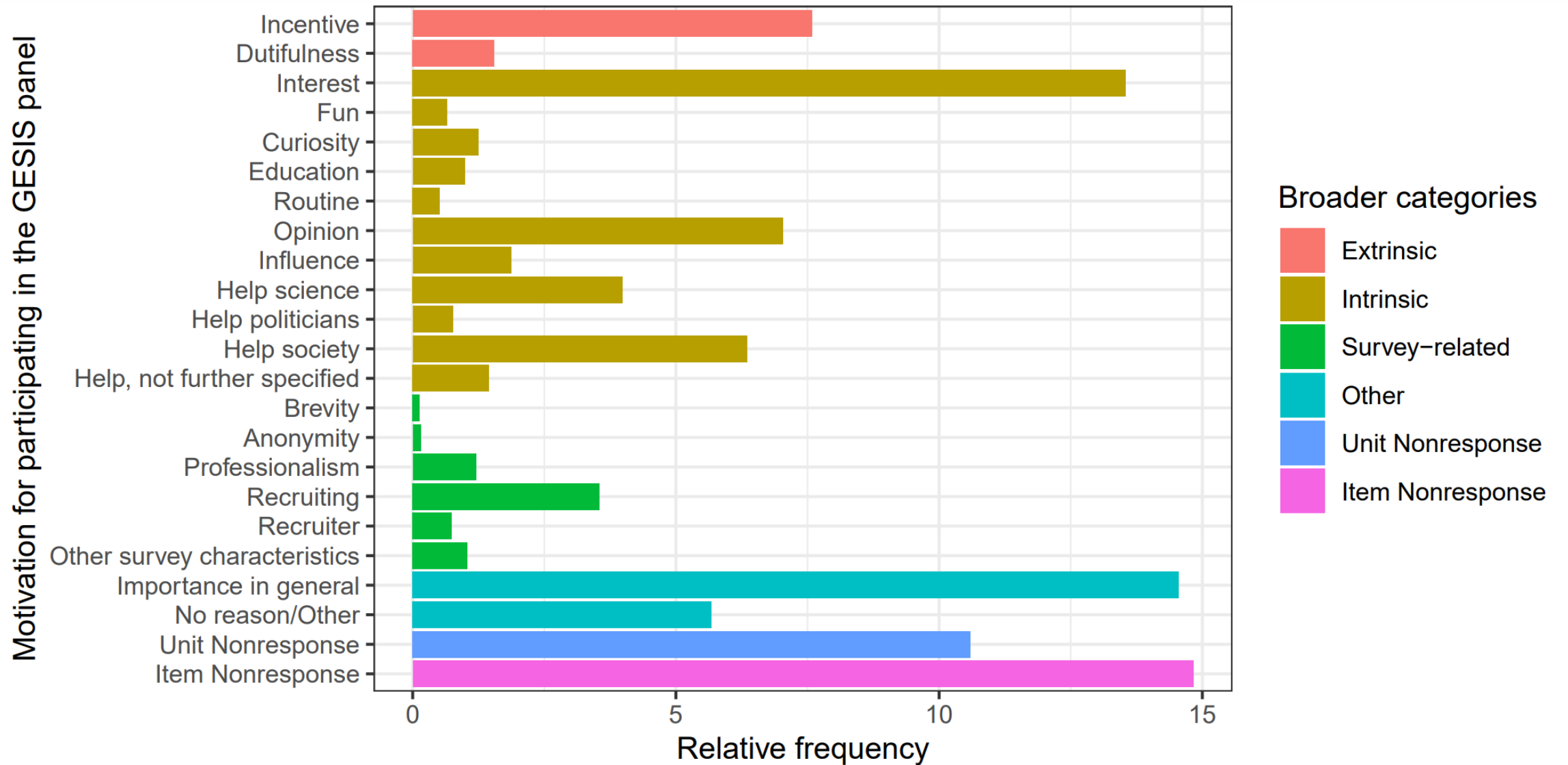
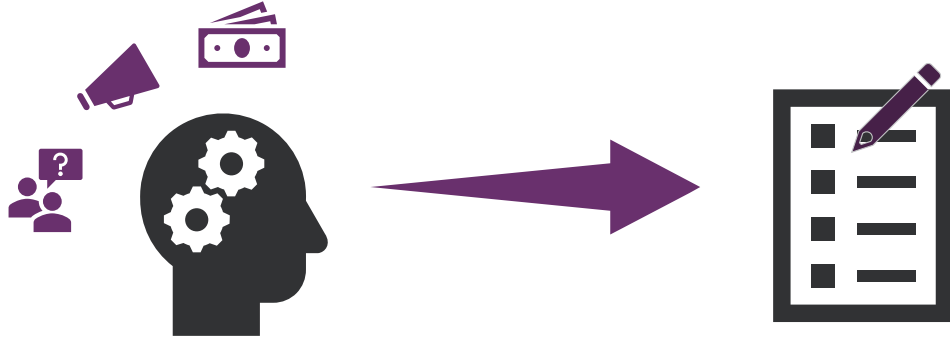


Figure: Most important reason in percent, wave “bf”, GESIS Panel.



# Data Quality



Rational choice theories establish a connection between the motivation to participate and participation or data quality

We examined the relationship between motivation to participate and data quality, e.g. *panel dropout*, *item nonresponse*, *straight-lining*

- Number of reasons are positively correlated with data quality
- Little differences between different types of motivations

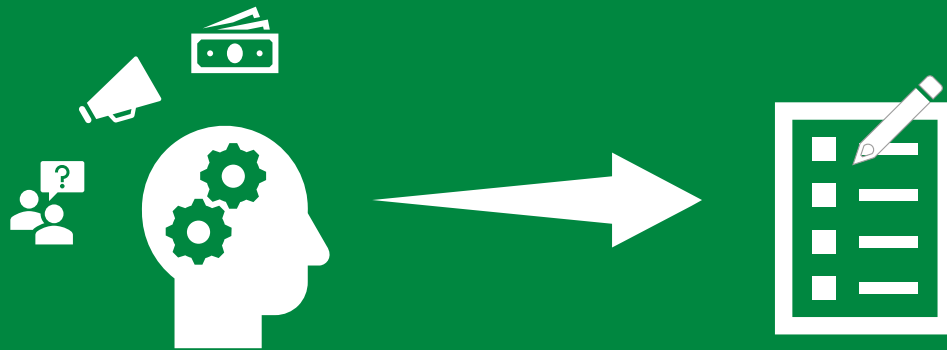




LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN



Thank you for your attention!



Email: [C.Haensch@lmu.de](mailto:C.Haensch@lmu.de) Twitter: [@CarolinaHaensch](https://twitter.com/CarolinaHaensch)

Preprint: <https://osf.io/preprints/socarxiv/sq2tm/>

