gesis Leibniz-Institut für Sozialwissenschaften

Using Predictive Modeling to Identify Panel Dropouts

Challenge:

- Panel studies suffer from attrition due to nonresponse, which may lead to a substantial loss in data quality.
- It is of utmost importance to predict which panelists are at risk of panel droppout.
- Once these at-risk panelists have been identified researchers can take appropriate measures to prevent panelists from dropping out.

Previous research:

Recent research emphasizes the importance of paradata in explaining attrition. It also make a strong argument for utilizing statistical learning algorithms to predict nonresponse.

Research questions:

1. Can we find evidence for the importance of paradata?
2. Does adding panel management information increase the performance of prediction?
3. How well perform statistical learning techniques, esp. ensemble methods (Lasso, conditional trees, random forest, gradient boosting)?

Data:

The GESIS Panel is a German probability-based mixed-mode (web-/mail-based) access panel ($n \approx 4,700$).

- *Outcome:* Nonresponse at wave ed (8/2017)
- *Predictors:*

  Socio-Demographics (SD), Health (HLTH), Survey attitude (SA), Paradata (PARA:Latency: Response latency [days]), Survey evaluation (EVAL), Panel management.

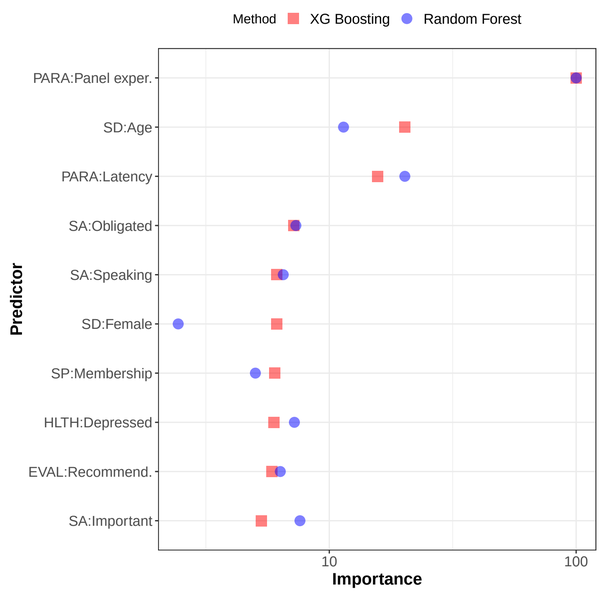
Methods:

Parametric methods (Logit Model, Lasso): $\boldsymbol{X}$ dependent on specification; linearity, additivity $\rightarrow$ Causation

Tree-based methods (CTrees, Random Forests, Gradient Boosting): „Built-in“ feature selection; no predefined functional form, complex interactions possible $\rightarrow$ Prediction

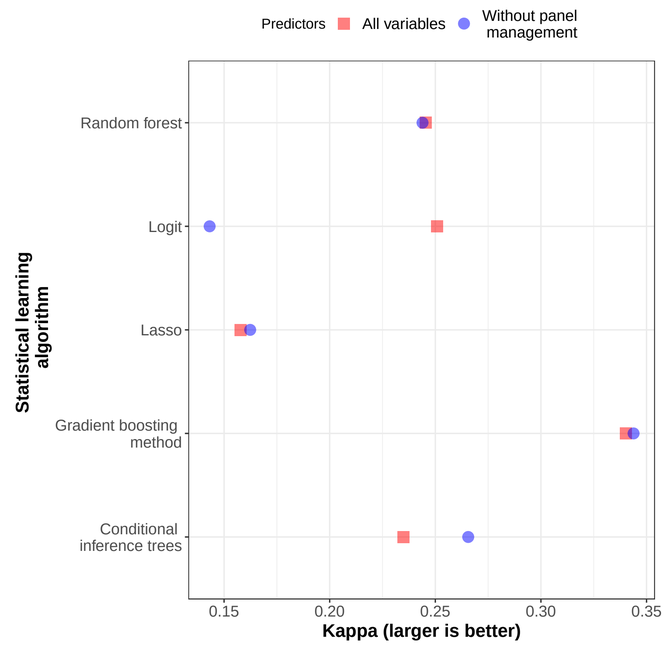
Results (RQ1)

Evidence - importance paradata



Results (RQ2 & RQ3)

Predictive impact of panel management information by various statistical learning techniques



Conclusion

- Paradata, survey evaluation & attitudes are important for predicting nonresponse.
- Including panel management information does not increase the Kappa measure (or accurancy) substantially.
- (Extreme) Gradient boosting and Random forest show best performance.

Kontakt: Jan-Philipp Kolb, Bernd Weiß (GESIS), Christoph Kern (Uni Mannheim)

Mitglied der Leibniz Gemeinschaft