CARMA 2026

8th Int. Conf. on Advanced Research Methods and Analytics
1-3 July 2026, Valencia, Spain
DOI: https://doi.org/10.4995/CARMA2026.2026.21847

Who really uses VAAs – Evidence of Diverging Biases Between Donor and Non-Donor Users

Lukas Schütte[1]

[1]University of Münster, Institute of Political Science, Germany.

How to cite: Lukas Schütte (2026). Who really uses VAAs – Evidence of Diverging Biases Between Donor and Non-Donor Users. In: 8th International Conference on Advanced Research Methods and Analytics (CARMA 2026). Valencia, 1-3 July 2026. https://doi.org/10.4995/CARMA2026.2026.21847

Abstract

This paper examines a neglected source of bias in Voting Advice Application (VAA) research: optional context-data donation. While prior work shows that VAA users are typically younger, more educated, and more politically interested than the general electorate, this study argues that a second selection layer arises when only some users complete optional metadata items. Using labelled Voto data (30,779 users) from 22 cities and unlabelled issue-response Voto data (262,034 users) from eight German city deployments, the paper trains tuned CatBoost models to predict party choice and left–right block affiliation from nine shared issue items. Results show substantial differences between donor and non-donor users indicating a much stronger right-leaning share in the unlabelled population than in the labelled donor sample. The findings suggest that donation nonresponse is not random and has direct implications for VAA inference, weighting, and data-release practices.

Keywords: *Voting Advice Applications; Supervised Machine Learning; Spatial Voting*

1. Introduction

Voting Advice Applications (VAAs) are widely used tools for political orientation and have become an important data source for electoral research. Classic VAA studies consistently show that users are not representative of the electorate. They are typically younger, more educated, more politically interested, and in many contexts more left leaning than the broader population. At the same time, VAAs do not only measure political attitudes; they also often include optional item-batteries asking for contextual or socio-demographic information. In most VAAs, only a subset of users completes these optional metadata batteries, while a much larger group provides only issue responses that are used to estimate party closeness. This raises an additional inferential problem: even within the VAA user base, the observable donor sample may differ systematically from non-donors leading to a second layer bias.

Article in Press

This work is licensed under a Creative Commons License CC BY-NC-SA 4.0
Editorial Universitat Politècnica de València

This paper addresses three questions. **H1**: users who do not provide context differ systematically from donors, with a tendency toward more right-leaning profiles. **H2**: bias in unlabeled published VAA responses can differ from the classic VAA user biases described in the literature. **H3**: Nonresponse to optional items is not random and may affect substantive inference. Empirically, the paper compares donor and non-donor users using predictive modeling trained on labeled VAA data and applied to unlabeled responses to estimate party/ideological preference. The results have practical implications for VAA deployment design, data donation prompts, and the public release and interpretation of VAA response data.

2. Data

2.1. VAAs as issue-space data

Voting Advice Applications (VAAs) collect users' positions on a set of policy questions and map these responses to party alternatives in a shared issue space. This design reflects the logic of spatial issue voting, where electoral choice is modeled as a function of proximity between voter and party positions (Garzia & Marschall, 2019; Louwerse & Rosema, 2014). For the present study, this provides the theoretical basis for using VAA response profiles to infer party and ideological alignment from shared issue items. The underlying logic is also consistent with spatial and rational-choice models of voting, in which party utility declines when policy distance increases (Downs, 1957).

At the same time, VAA data are affected by well-known selection effects, as users are typically younger, more educated, and more politically interested than the general electorate (Pianzola, 2014). This paper focuses on an additional and less studied selection stage, namely the optional data donation. In many VAA settings, only a subset of users provide contextual sociodemographic information, which creates a labeled subsample that may differ systematically from the much larger pool of users with issue responses alone (Kersting & Schütte, 2026; Suhay et al., 2020). The central question is therefore not only whether VAA users are biased, but whether the observable donor sample is biased relative to the full VAA user population.

2.2 Data source and samples

The analysis combines labeled VAA data from a former combined dataset added to our projects data donations (Stecker et al., 2025). The labelled corpus contains 30,779 observations from 22 German cities utilized for model training, testing and validation. This dataset includes users' responses to a shared set of nine policy items, and the users vote intention as the target feature required for supervised learning. Because only nine policy items are shared across all city deployments, the predictive task is intentionally restricted to a compact common issue space.

Article in Press

This design maximizes comparability across cities but limits the amount of information available to the model, especially for fine-grained multiclass party prediction.

The unlabeled dataset contains 262,034 VAA responses from our Voto deployments in Köln, Münster, Paderborn, Bielefeld, Essen, Oberhausen, Duisburg and Bochum. For these users, issue responses are available, but optional contextual donation variables are missing, making it a perfect setup for prediction-based evaluation.

A crucial design feature is that 9 issue items are identical across all cities. These shared items form the common analytical issue-space for model training and transfer to the unlabeled sample.

2.3 Items and operationalization

The operationalization follows the same theoretical logic as the VAA itself: party choice is treated as a function of users' positions on policy statements. Specifically, the model uses the nine shared issue items as predictors, where each item captures a policy stance relevant to party competition. Together, these items define a compact issue-space profile for each user.

3. Methodology

To estimate the political composition of the unlabeled VAA population, two supervised classification models were trained on the labeled VAA sample and then applied to the unlabeled cases. CatBoost was used in both cases because it performs well on structured survey-style data and supports robust class-weighting for imbalanced targets.

3.1 Predictive setup and Inference strategy

The first model estimates party choice as a multiclass outcome. To reduce noise and improve class separability, the training target was restricted to five parties: AfD, CDU, SPD, Die Grünen, and Die Linke. BSW, FDP, and small parties were excluded from the multiclass pipeline because they reduced model stability: BSW and FDP had too few observations for reliable training, while the grouped small-party categories were substantively too heterogeneous for a consistent party-level target. Party labels were encoded numerically, and the dataset was split using a stratified procedure into training, test, and validation sets. Hyperparameters were tuned on the training set using CatBoost's randomized search with 5-fold CV (30 sampled parameter combinations). The search varied tree depth, learning rate, regularization, feature subsampling, number of iterations, and Bayesian bagging temperature. The final multiclass CatBoost model was then trained with the best parameter set and early stopping based on validation performance. The second model estimates a binary ideological outcome (left vs. right) as a more aggregated specification. In contrast to the multiclass model, this binary target includes FDP and the grouped small parties, because these categories can be assigned to broader ideological blocks. The binary coding maps SPD, Die Grünen, Die Linke and progressive small parties to the left

Article in Press

category, and AfD, CDU, FDP, and small conservative parties to the right category. It uses the same 9 issue predictors and the same stratified splitting logic. Hyperparameters were again tuned via CatBoost randomized search with 5-fold CV (30 iterations), using a binary classification setup (log loss objective, AUC as tuning metric). As in the multiclass pipeline, the final model was trained with the best parameter set and early stopping on the validation set. After evaluation, both tuned models were applied to the unlabeled VAA users to generate predicted political distributions. The multiclass model provides a party-level estimate for the five-party space, while the binary model provides a broader left–right estimate. Comparing these predicted distributions to the observed labeled sample allows an assessment of whether donor and non-donor users differ systematically in their issue-based political composition.

3.2 Model Performance

Table 1. Evaluation Metrics Multi-Level Tuned CatBoost

	precision	***recall***	***F-1***	***Support (test)***
AfD	0.67	0.78	0.72	644
CDU	0.62	0.55	0.58	1022
Die Grünen	0.74	0.5	0.6	1976
Die Linke	0.33	0.57	0.42	691
SPD	0.35	0.4	0.37	854
Macro accuracy			0.54	5187

Table 2: Evaluation Metrics Binary Tuned CatBoost

	precision	***recall***	***F-1***	***Support (test)***
Left-Branch	0.89	0.88	0.89	3673
Right-Branch	0.82	0.84	0.83	2402
Macro accuracy			0.86	6075

The tuned multiclass CatBoost achieves 0.54 accuracy on the held-out test set (N=5,187) and shows uneven performance across parties. It performs best for AfD and reasonably for CDU and Die Grünen. Performance is weaker for SPD and Die Linke, indicating that party-level distinctions are harder to recover for these classes. Importantly, inspection of misclassifications suggests that many errors occur between substantively adjacent parties, rather than across ideologically distant categories. This makes the multiclass errors less problematic for the paper's broader inferential purpose. The tuned binary CatBoost performs substantially better, reaching 0.86 accuracy on the test set (N=6,075). The moderate multiclass accuracy should be interpreted considering the restricted nine-item feature space. These items capture broad

Article in Press

ideological structure sufficiently well, as indicated by the stronger binary model performance, but they provide only limited information for distinguishing between ideologically adjacent parties such as SPD, Die Grünen, and Die Linke. Classification quality is high for both branches, with left at F1=0.89 and right at F1=0.83. The balanced performance across both categories indicates that the issue-item predictors capture the broader ideological cleavage very well. Substantively, this supports the interpretation that many multiclass errors reflect within-bloc ambiguity rather than failure to recover underlying political orientation.

4. Results & Discussion

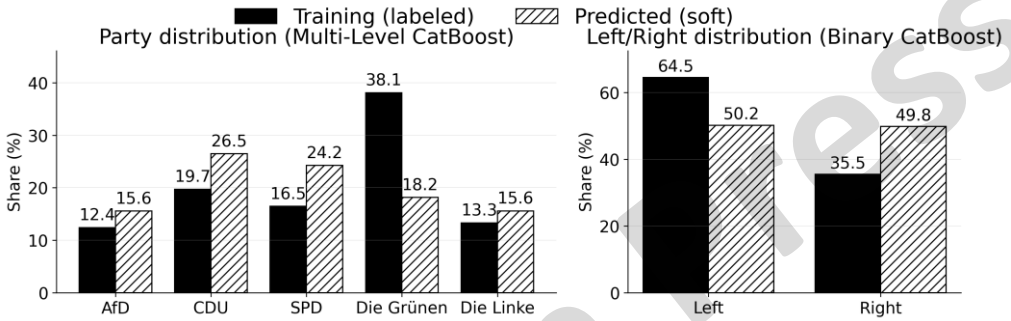


Figure 1: Share of Predicted vs Labeled Data

Figure 1, displaying soft prediction shares (each response contributes to each party relative to its predicted probability), shows a clear gap between the labeled donor sample and the predicted distribution in the unlabeled VAA sample. In the multiclass model, the unlabeled group contains markedly higher predicted shares for CDU, and SPD, whilst Die Grünen are substantially lower. AfD and Die Linke are also somewhat higher in the unlabeled group. Even with known limitations of the multi-class model, the direction shift is consistent with expected differentiations. This pattern is reinforced by the binary model, which is much more reliable in its predictive power. In the labeled sample, the left branch dominated, but in the unlabeled sample the distribution is nearly balanced, with a string increase in the predicted right branch. Given the binary model's high test performance, this strongly supports H1 and H3: non-donors differ systematically from donors, and nonresponse to optional context items is not random. Substantially, this also supports H2.

The analysis cannot directly estimate socio-demographic differences between donors and non-donors, such as age, education, or gender, because these contextual variables are unavailable for non-donors. The contribution is therefore narrower: it evaluates whether donor and non-donor users differ in their issue-space profiles and predicted political orientation. Future VAA designs might collect minimal socio-demographic indicators earlier or independently of

Article in Press

optional post-recommendation donation batteries to test whether similar nonresponse patterns also occur for socio-demographic characteristics.

What the results show is that the bias in analyzable VAA datasets may differ from the VAA biases described in literature. The issue means not only that VAA users differ from the electorate, but that context donors differ from non-donors within the VAA framework. This matters for inference, weighting strategies when analyzing unlabeled data and assuming similar biases and claims about who uses VAAs in general. A plausible explanation is design-related nonresponse. If the context battery appears after the party recommendations, users may leave the tool before reaching it, and this dropout may vary by ideological orientation, engagement style or age. The results therefore suggest a practical design implication. Data donation should be made more visible and ideally requested earlier and transparently at the start. Future research should test whether the same nonrandom donation pattern also affects other context variables such as age, and education. Beyond methodology, these differences may also help VAA providers to better understand and strategically address a broader user audience that is currently underrepresented in VAA settings and to market their tool beyond common target groups.

References

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, *65*(2), 135–150. https://doi.org/10.1086/257897

Garzia, D., & Marschall, S. (2019). Voting Advice Applications. In *Oxford Research Encyclopedia of Politics*. https://doi.org/10.1093/acrefore/9780190228637.013.620

Kersting, N., & Schütte, L. (2026). Kongruenz und Divergenz zwischen Wählerschaft und Parteien bei der Bundestagswahl 2025. *Zeitschrift Für Parlamentsfragen*, *57*(4), 733–749.

Louwerse, T., & Rosema, M. (2014). The design effects of voting advice applications: Comparing methods of calculating matches. *Acta Politica*, *49*(3), 286–312. https://doi.org/10.1057/ap.2013.30

Pianzola, J. (2014). Selection biases in Voting Advice Application research. *Electoral Studies*, *36*, 272–280. https://doi.org/10.1016/j.electstud.2014.04.012

Stecker, C., Dieing, T. I., Oblong, J., Strobel, S., Braun, D., Deiss-Helbig, E., Gessler, T., Hein, M., Höhne, B., Kueppers, A., Linhart, E., Müller, J., Pravemann, S., Reiser, M., Steiner, N. D., Thomeczek, J. P., Völkl, K., Weisskircher, M., Wenzelburger, G., & Zissel, P. (2025). *Voting Advice Applications for Local Elections* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/PAH9O1

Suhay, E., Grofman, B., & Trechsel, H. (2020). Voting advice applications: The power of self-persuasion. In *The Oxford handbook of electoral persuasion* (pp. 925–045). Oxford University Press. https://cadmus.eui.eu/entities/publication/eef61a60-f654-5277-99c8-911b0fff9f40

Article in Press