A Computational Decision-Tree Approach to Inform Post-Conviction Intake Decisions

Kalina Kostyszyn, Carl J. Wiedemann, Rosa Bermejo, Amie Paige,
Kristen W. Kalb-DellaRatta, & Susan E. Brennan

Overview

We explore data from the National Registry of Exonerations (NRE; 4/26/2023, N = 3,284 exonerations) to inform such decisions, using patterns of features associated with successful prior cases. We first reproduce Berube et al. (2023)'s latent class analysis, identifying four underlying categories across cases. We then use decision trees (WEKA, Frank et al., 2013) to decompose complex patterns of data into ordered flows of variables, with the potential to guide intermediate steps that could be tailored to the particular organization's limitations, areas of expertise, and resources.

National Registry of Exoneration

The *National Registry of Exonerations* (NRE) collects federal and state exonerations in the United State from 1989 onward, including limited information prior to that. Data collected includes:

- Age
- Race
- State, County of Crime
- 18 Case Tags
- 11 Official Misconduct Tags
- Crime
- Sentence Length
- Year Convicted
- Year Exonerated
- 6 Canonical Factors of Wrongful Conviction

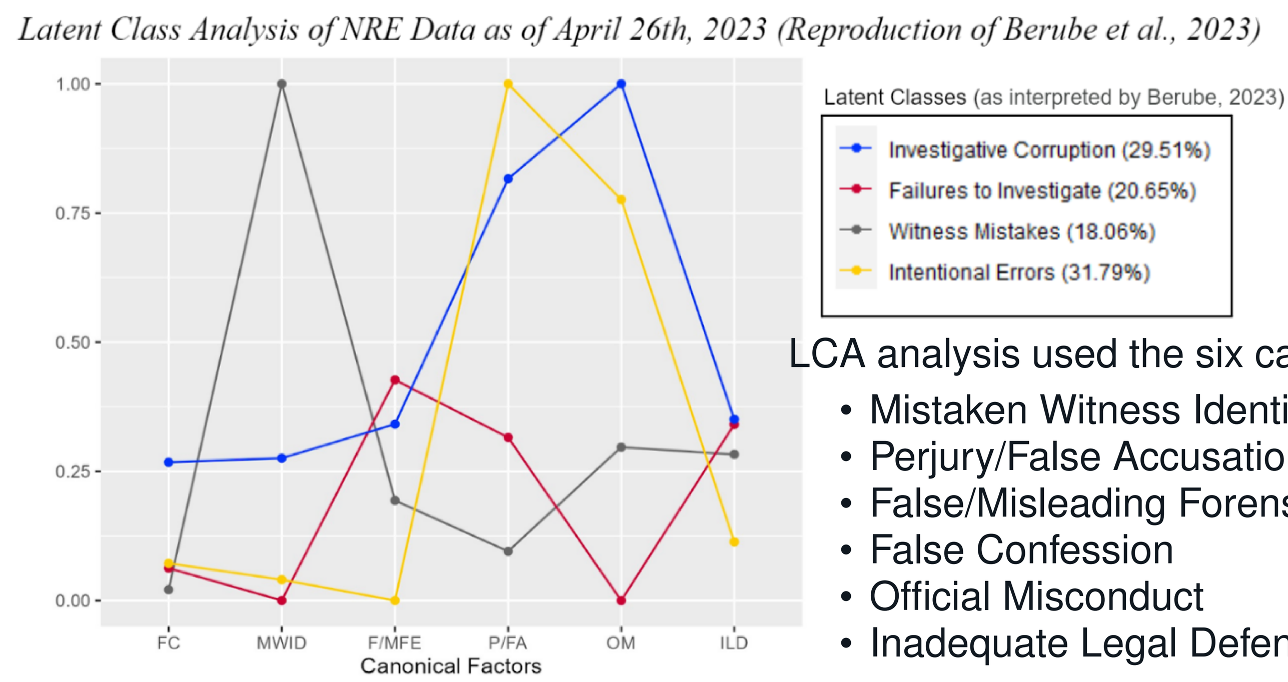
Goals

How can decision trees organize exoneration data in a way that's transparent and interpretable for innocence organization staff?

How can interaction between decision trees and latent class analyses clarify commonalities within classes?

How can this analysis be modified to suit the needs of innocence organization staff?

Latent Class Analysis



LCA analysis used the six canonical factors:

- Mistaken Witness Identity
- Perjury/False Accusation
- False/Misleading Forensic Evidence
- False Confession
- Official Misconduct
- Inadequate Legal Defense

Latent Class Analyses (LCAs) are informative for datasets such as the NRE, as they identify *latent* (i.e., not directly observable) subgroups within populations (McCutcheon, 2002).

Decision Tree Results

Model	Precision	Recall	F-Score
6-factor, cross validated	1.000	1.000	1.000
6-factor, 75-25	0.999	0.999	0.999
Extended, cross validated	0.722	0.721	0.720
Extended, 75-25	0.737	0.737	0.737

Precision and recall scores suggest that the decision tree can suitably replicate the latent classes, using either *the canonical factors* or the *set of case and misconduct tags*.

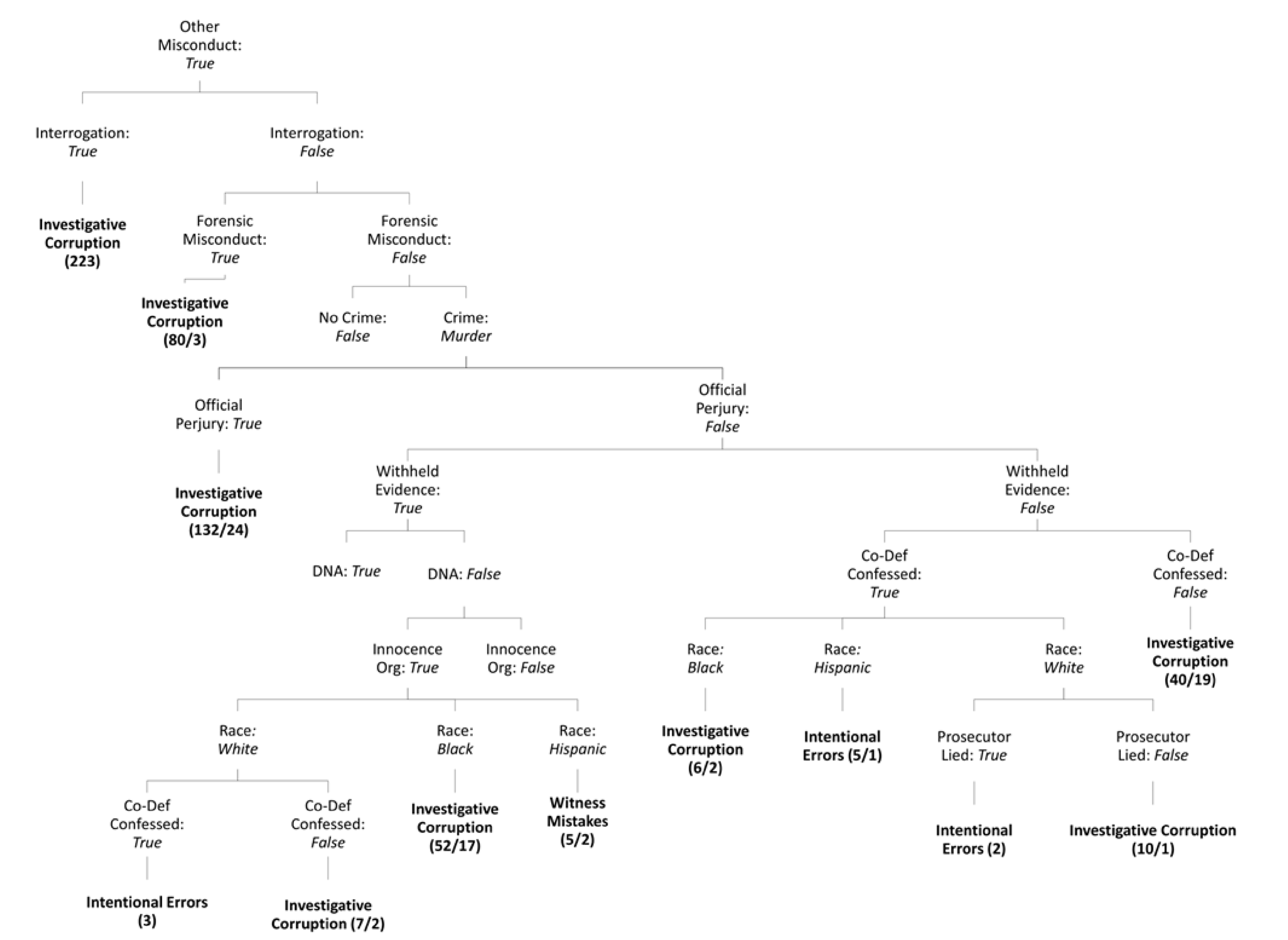
	Classed FtI	Classed IE	Classed IC	Classed WM
True FtI	131	16	0	12
True IE	19	173	**56**	11
True IC	0	**53**	176	12
True WM	10	11	15	124

Confusion matrices can inform about patterns in 'misclassifications' (or, disagreements in classification between the LCA and decision tree). For example, the two values in bold exhibit the highest misclassification pattern in the models we generated, with the consistent pattern that the to right (True Intentional Error) always is higher than the bottom left (True Investigative Corruption).

Model	Precision	Recall	F-Score
No DNA, no-crime, cross validated	0.714	0.711	0.711
No DNA, no-crime, 75-25	0.723	0.720	0.719
No IO, CIU, cross validated	0.718	0.717	0.716
No IO, CIU, 75-25	0.750	0.750	0.749
Added state, cross validated	0.722	0.719	0.719
Added state, 75-25	0.731	0.730	0.730

We then made three modifications to the data. The modifications were based on excluding 'current knowledge' (whether DNA was used in exoneration and whether it was determined that no crime had actually occurred), on excluding interventions (involvement of an innocence organization or conviction integrity unit), and including state information, to see if location of the crime held informational weight (suggesting potential for systemic issues at that location, etc.).

Example tree segment



Future directions

- Adding election data per county to infer effect of political alignment on wrongful convictions?
- Specific impacts on non-native English speakers
- Collaboration and input from Innocence organizations so future development is in line with staff needs

Selected References

Berube, R., Wilford, M. M., Redlich A. D., & Wang, Y. (2023). Identifying patterns across the six canonical factors underlying wrongful convictions. The Wrongful Conviction Law Review, 3(3), 166-195. https://doi.org/10.29173/wclawr82

Frank, Eibe, Mark A. Hall, and Ian H. Witten. 2016. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques" (4th ed.). Morgan Kaufman.

McCutcheon, A. (2002). Basic Concepts and Procedures in Single- and Multiple-Group Latent Class Analysis. In J. Hagenaars & A. McCutcheon (Eds.), Applied Latent Class Analysis (pp. 56-86). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511499531.003

This material is based upon work supported by NSF under Grant NRT-HDR 2125295. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Stony Brook University

iACS Institute for Advanced Computational Science

Detecting and Addressing BIAS in Data, Humans, and Institutions