

Purpose

The purpose of this project is to build a machine learning model that predicts insurance status based on social determinants of health selected as input variables

Background

- Social determinants of health (SODH) are factors that influence an individual's health status.
- SODH inequalities often negatively impact health status.
- LGBTQ+ individuals are more likely than their heterosexual counterparts to be unemployed and uninsured in early adulthood (Charlton, Gordon, Reisner, 2018).

Methodology

- Our data was accessed from the National Center for Health Statistics' 2019 National Health Interview Survey.
- Participants (n=590) were selected based on gender identity ("female") and sexual orientation ("lesbian", "bisexual", "something else").
- SODH: Education access/quality, economic stability, social and community context, health care access/quality, and neighborhood and built environment.
- A Conditional Inference Tree was developed using the "partykit" package in the statistical software "R."

Results

- Roughly 20% of individuals who indicated they had trouble paying medical bills in the last twelve months were not insured ($p < 0.001$).
- Individuals who reported a shorter length of time since their last doctor visit (i.e., within the last year) were more likely to be insured than those who reported a longer period since their last visit ($p = 0.001$).

Discussion

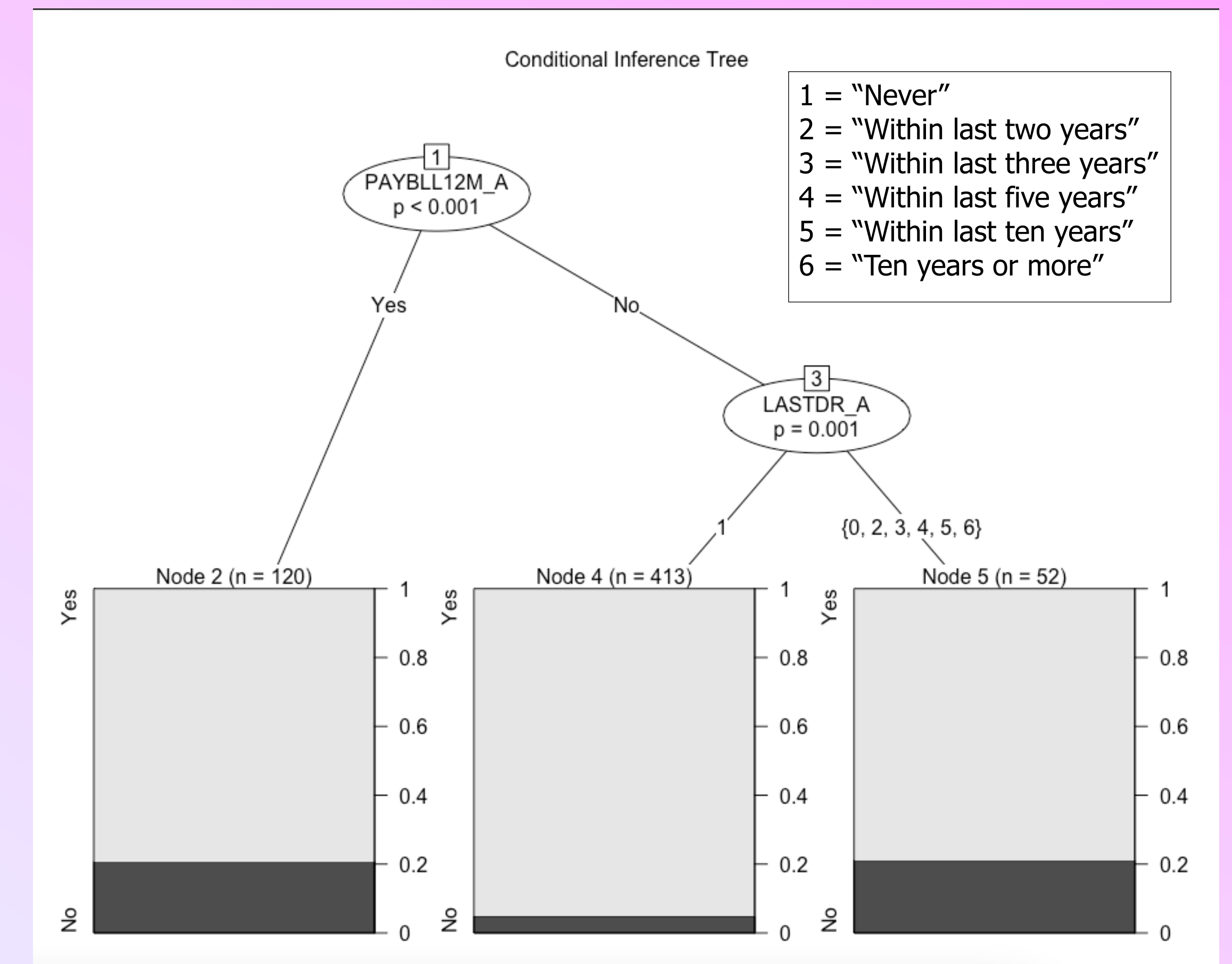
- We can use this information to identify points of contact for delivering Medicaid application assistance, information about low-cost healthcare providers, and other options that promote utilization of health care services at an affordable cost.

- Conditional inference trees (i.e., unbiased recursive partitioning) are a nonparametric approach to classification and regression tree analysis.
- The tree begins at the root node and splits into branches based on outcome and splitting criterion.
 - The terminal (leaf) nodes are the final outcomes.
- The minimum bucket size was set to 50, indicating that each terminal node decision must pertain to at least 50 individuals.

Analytic Approach

Conditional Inference Tree Algorithm

1. Test the global null hypothesis between independent and response variables and select the independent variable with the highest p-value with response variable.
2. Implement binary split on the selected input variable.
3. Recursively repeat steps 1 and 2 until a terminal decision is made.



References

Charlton, Brittany M, et al. "Sexual Orientation-Related Disparities in Employment, Health Insurance, Healthcare Access and Health-Related Quality of Life: a Cohort Study of US Male and Female Adolescents and Young Adults." *BMJ Open*, vol. 8, no. 6, 2018, doi:10.1136/bmjopen-2017-020418.

Healthy People 2030, U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Retrieved Apr. 9, 2021, from <https://health.gov/healthypeople/objectives-and-data/social-determinants-health>

Torsten Hothorn, Achim Zeileis (2015). partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research*, 16, 3905-3909. URL <https://jmlr.org/papers/v16/hothorn15a.htm>

Torsten Hothorn, Kurt Hornik and Achim Zeileis (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3), 651--674, DOI:10.1198/106186006X133933