

M244 Peer Review Feedback for Caroline & Eli

Boxiong and Yifei

Introduction and Data

- Nice introduction to the variables with clear explanation
- This is just a minor point - instead of saying “there were a lot of NA values”, try to specify how many were dropped and how many observations were left for analysis.

Methodology

- Explain in more detail how INTL_ECONOMICIMPACT is transformed from string to numerical values
- Shorten the summary table for datasets - maybe use Knitr for better formatting
- Add more clarification to the decision tree graph
- Adjust figure 2 to a stacked bar plot to illustrate the change of percentage over time
- Adjust plots' axes labels to fit the size of the figure

Results

- Add short descriptions of the tree's result in context can help better understanding of the result
- Could be better to put the true label vs. predicted label plot into Results section to correspond with the evaluation of model performance, and further interpret the matrix

Discussion

- Add a section discussing the impact of arbitrarily assigning Jan 1st to years
- Try to think of the limitation of using survey data to generalize about group behavior — the responses might be biased
- Very critical ethical concerns

Formatting

- Very nice formatting - all codes are commented so we can easily follow every step you take

General

- Some code chunks can be hidden for better readability, such as EDA visualizations, where only the plots need to be retained in the report while hiding the code that generates the plots
- Some code outputs can be formatted more neatly—for example, the # Print categories in order section and the probability array from `clf.predict_proba`
- All writings and codes are clear

- Haven't had access to the GitHub repository, but try to ensure all related datasets and files are properly included.

Final Considerations

- The introduction is thorough and well-structured, having strong motivation by referencing multiple studies centered on climate change. It also delivers a precise breakdown of all relevant variables and establishes logical assumptions for the research question.
- Question:
 - How might respondents' income or socioeconomic status confound the observed relationship between political ideology and willingness to pay, and could you incorporate these variables into your models to test their influence?
 - Given the large amount of missing data in certain waves and variables, have you explored multiple imputation or sensitivity analyses to assess whether your findings are robust to different missing-data assumptions?