Overview

The package is simple. Please run the do files and python script in the order explained before.

Data Availability

The data used in this study belong to the Latinobarometro Corporation and can be found and downloaded here:

https://www.latinobarometro.org/latContents.jsp

Instructions for Replicators

New users should follow these steps to run the package successfully:

- The folder called *DataIn* contains the raw data used for the analysis, please be sure to keep it in that folder. The name of the .dta is Latinobarometro_2023_Eng_Stata_v1_0
- In the folder named *Code* you will find the do file `'Latinobarometro 2023.do'` that cleans the raw data and prepares the variables for the analysis. Make sure you run this do file before any other steps. The do file will create a clean dataset in .dta format which will be saved in the folder *DataOut* with the name Latinobarometro_2023.dta
- The analysis is performed completely in Python, but to correctly run the decision trees algorithms, missing values need to be handled first. This is what the do file called `'Latinobarometro 2023 Prep for Python'` does. So, make sure you run it before proceeding to Python.
- Open Python and run the script named `'Gender and AI in civic participation 2024'`. You can run the whole script all at once and the outputs (the figures and decision trees) will be sequentially generated in the 'plots' tab in the console. The key package needed will be scikit-learn (we used version 1.2.2). Python version is 3.11

List of Exhibits

The provided Python code reproduces the following figures in the text. You will find the reference also in comments in the script. Please note that the figures can be exported in a limited number of formats so an editor should be happy with a PNG image as we cannot provide them in any other format.

- Figure 2. Baseline model: decision tree for citizen participation
- Figure 3. Feature importance for the baseline model for citizen participation
- Figure 4. Composite gender score in success path with varying scaling and shifting factors
- Figure 5. Partial dependence plot for the effect of education on working on a community problem
- Figure 6. Interaction plot: Effect of gender bias score and education levels on community problem involvement

Requirements

Computational and Software Requirements

You need 64 gb of RAM memory and 16 cores. You need Stata 18 and Python 3.11 and install the following packages in Python:

- pandas
- numpy
- seaborn
- matplotlib
- sklearn
- graphviz
- difflib
- contextlib

Memory and Runtime and Storage Requirements

Runs fast so don't worry.

Code Description

First run `'Latinobarometro 2023.do'`

Then Run `'Latinobarometro 2023 Prep for Python'`

Then open Python and run `'Gender and AI in civic participation 2024.py'`. Note that this script will use as data frame the file `'Latinobarometro_2023_no_missing.dta'`. Also note that the figures are produced in the console, they are not stored in any folder.

Folder Structure

```
DataIn
   ├── Latinobarometro_2023_Eng_Stata_v1_0.dta
   └──
Code
   ├── Latinobarometro 2023.do
   ├── Latinobarometro 2023 Prep for Python.do
   └── Gender and AI in civic participation 2024.do
DataOut
   ├── Latinobarometro_2023
   └── Latinobarometro_2023_no_missing
```