RDS Final Project Presentation

Oviya Adhan and Mina Mohammadi
NYU
May 2021

Background and Purpose:

- The **ADS' goal** is to distribute aid more efficiently taking into account factors that have previously been ignored like specific household attributes like material of their walls, ceilings and other related assets. These attributes better help the ADS and thus the policy makers to better directly target most at need groups and provide social goods responsibly.

- **Stakeholders** in this ADS that would benefit include the Costa Rican government and poor families in need.

Trade-Offs

- The ADS here seems to have only **one goal** of understanding the most at need and at risk groups in poverty. The goal is to tune the model to the right variables to create the most accurate system to target the right people in poverty.

- One **potential trade off** in this ADS because determining that one variable is more highly weighted than another variable, its real world consequence is that some families may miss out on aid because the algorithm makes their poverty indicator a lower weighted variable.
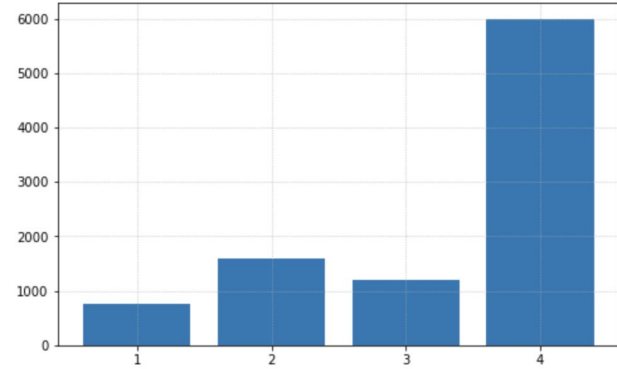
Input

- Our data has 142 input variables, we used the .dtypes function to illustrate the **data types** into our code.

- There were 5 input variables with **missing values** listed here:
  - v18q1 (number of tablets household owns)
  - rez_esc (years behind in school)
  - meaneduc (average number of years of education)
  - v2a1 (monthly rent payment)
 - SQBmeaned (squared mean years of education of adults in the household).

Output

- The **output** is an ordinal variable the level of impoverishment of a household.
- Target - the target is an ordinal variable indicating groups of income levels.
  1 = extreme poverty
  2 = moderate poverty
  3 = vulnerable households
  4 = non vulnerable households
- In other words, the lower the score, the more impoverished the household, the more likely IDB would provide aid to that household.

The model

- The input of the ADS includes various variables including tangible features of the houses of the households, gender ratios of the households, education levels, etc
- The ADS was validated with these methods:

  -prediction made with model

  -accuracy checked (about ~91%)

  -created a confusion matrix heat map
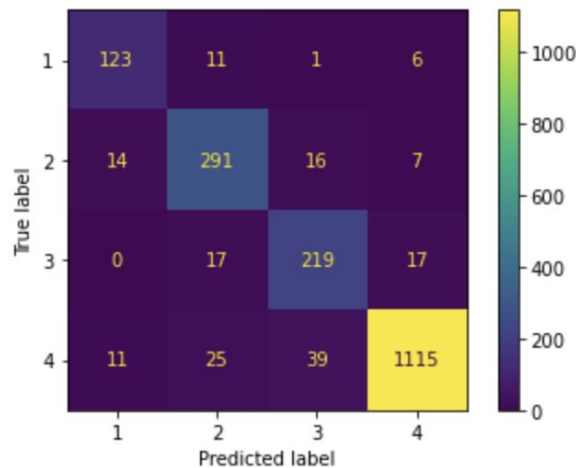
High-level information

Input
- Various variables

Output
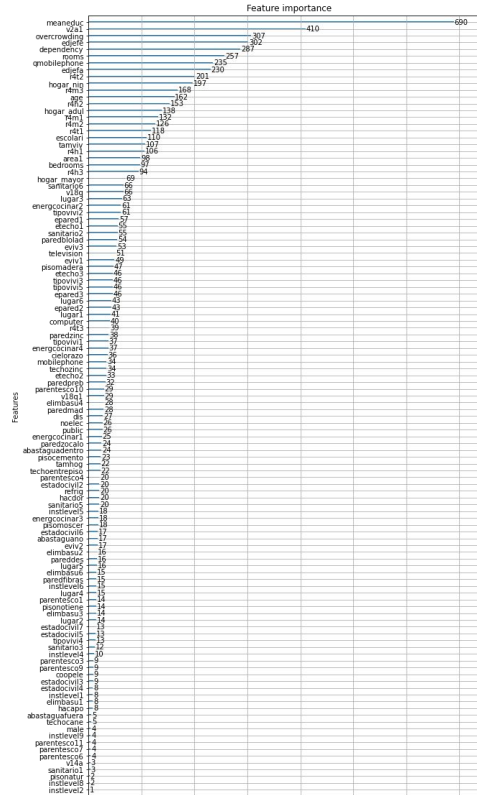- Impoverishment Level (Ordinal variable)

Gradient Boosted Decision Trees (XGB)
- Gradient Boosted regression to evaluate features importance

Outcomes

Effectiveness (Accuracy)

- Overall model accuracy = 77.2489539748954
- Females = 76.6260162601626
- Males = 77.90948275862068

Top 5 Features:

1. Meaneduc
2. V2a1 (monthly rent payment)
3. Overcrowding
4. Dependency
5. Rooms

Summary

- We would argue that this data is appropriate for the ADS when compared to traditional econometrics methods (more targeted indicators)
- The implementation of the ADS is accurate and fair
  - We would however say that the robustness of the model could be improved
- We would be comfortable deploying this ADS into the public sector in Costa Rica
  - it better micro targets specific familial need based issues to better target aid