Algorithmic Targeting of Social Policies: Accuracy & Fairness

***This work is a collaboration between research institutions of the USA and Mexico. The two first authors are Latin American and the senior author is from the USA.**

1 Introduction

Targeted social policies are the main strategy for poverty alleviation across the developing world [7, 3]. Moreover, since two decades ago, algorithmic rules underlie the targeting decisions of a large fraction of social policies in the developing world, commonly in the form of poverty estimation models (called *proxy means tests*) [8, 7, 6]. In many countries, governments and agencies such as USAID [5] use such models to target programs that touch every corner of social development: from cash transfer programs (CTs), to scholarship programs, subsidized health care systems, targeted housing, childcare, and others. These algorithms directly influence the lives of hundreds of millions of individuals in the developing world [8], and are, thus, among the most important algorithms in operation today. We show that a shift towards the use of machine learning (ML) in poverty-based targeting can substantially increase accuracy, extending the coverage of the poor by nearly a million people in two Latin American countries. In both countries, ML methods reduced predictive errors for every population subgroup, compared to status quo methods, as well as narrowed the disparities across subgroups. To the best of our knowledge, this is the first time that algorithmic rules of this sort are audited for potential exclusion disparities across population subgroups.

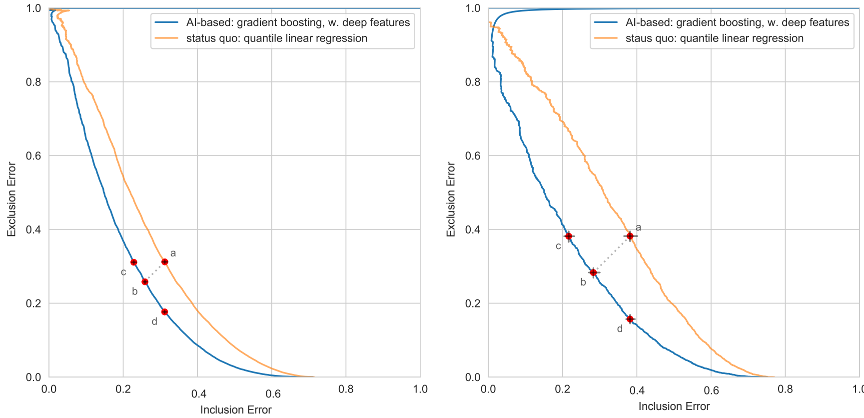
2 Empirical Evaluation

Data. We obtain the income ground truth and relevant household information, such as individual's education levels and the household's assets, from publicly available household surveys from Colombia and Costa Rica. This is the data on which current targeting systems were trained and evaluated [8, 1].

Algorithms. *Status quo:* The predictors currently used for income-based targeting of social policies stem mainly from econometric methods [5, 8]. In particular, quantile linear regressions [9] are commonly used. We will refer to quantile linear regressions as the *status quo* methodology. *ML-based:* We preprocessed the survey data to generate three types of features. First, *expert features*, crafted by human experts. Second, *statistical features*, including means, modes and entropies for all individual-level variables of household members. Lastly, *deep features*, generated by a recursive neural network that condenses information of the individual-level features into a one-dimensional encoding—a technique akin to multiple instance learning (MIL) [4]. The best performing algorithm was a gradient boosting classifier trained on the three feature sets concatenated. Meta-parameters were determined via k-fold cross-validation. We refer to this as the *ML-based method*.

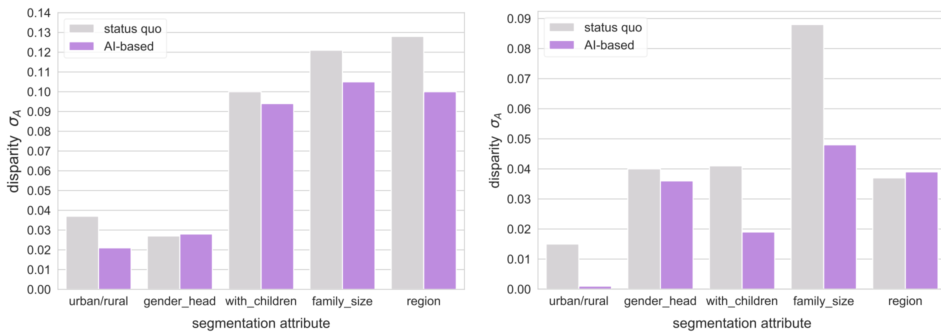
Metrics. The two key metrics used by institutions in the social sector are the exclusion and inclusion errors [8, 7]. The ***exclusion error*** measures the percentage of poor households incorrectly classified as non-poor, denoted by $\epsilon_{\text{exc}} = \frac{\text{FN}}{\text{TP+FN}}$; while the ***inclusion error*** measures the percentage of non-poor households incorrectly classified as poor, denoted by $\epsilon_{\text{inc}} = \frac{\text{FP}}{\text{TP+FP}}$. TP, FP, TN, FN correspond respectively to true positives, false positives, true negatives, and false negatives. We also compute the ***exclusion-inclusion curve*** (EIC), which maps the entire space of targeting rules. At the subgroup level, we define the ***subgroup disparity*** σ_A as the standard deviation of the exclusion errors across subgroups, i.e., $\sigma_A = \sqrt{\sum_{g \in A} (e_g - \bar{e})^2}$, where A is a segmentation attribute such as *family size*, and $g \in A$ are the subgroups that it defines.

Results. *Overall improvements:* Figures 1a and 1b present the performance comparison for Colombia and Costa Rica. A social program switching from $\boldsymbol{a}$ to $\boldsymbol{b}$ would reduce both its exclusion and inclusion errors by 17.3% in Colombia, and by 25.9% in Costa Rica. This reduction in errors in both countries means that 840,000 people in poverty, previously misclassified, would be now correctly included in the countries' set of social protection policies. *Subgroup-specific improvements:* Table 1



(a) Colombia (b) Costa Rica

Figure 1: **Exclusion error versus inclusion error curve.**



(a) Colombia (b) Costa Rica

Figure 2: The ML-based method substantially reduces subgroup disparities.

provides an overview of the relative reduction in group-specific errors for all combinations of the two countries and five segmentation attributes: urban/rural, gender of the head, with children, family size, and region(*region*'s overview can be found in A). We find that the ML-based method reduced the errors for each subgroup across the five segmentation attributes and two countries. *Disparities across subgroups:* The ML-based method also balances error disparities across subgroups. Figure 2 shows the subgroup disparities σ_A for the status quo method and the ML-based method. Although disparities persist when using the ML-based method, σ_A disparities were reduced in both countries and for most segmentation attributes.

Table 1: Relative decrease in exclusion errors obtained by the ML-based versus the status quo.

	Urban		Gender		Children		Family Size					
	T	F	M	F	w/	w/o	1	2	3	4	5	6
Colombia	18.6	10.6	17.1	17.5	15.7	19.3	26.1	16.1	16.7	8.5	13.1	13.2
Costa Rica	31.5	23.5	26.4	24.2	23.3	34.9	46.4	28.3	31.4	24.4	15.9	14.0

3 Conclusions

This work provides three key results. First, ML-based targeting reduced exclusion and inclusion errors, thus extending the coverage of social programs by nearly a million people in both Colombia and Costa Rica without increasing expenditure. Secondly, ML-based methods reduced predictive errors for every population subgroup in both countries. Finally, the ML-based method narrowed the disparities across the subgroups when compared to the status quo. These results ease a transition from the status quo to an ML-based targeting system. Despite these advantages, disparities persist after introducing ML-based methods. Forthcoming work provides an answer to this issue by means of an interactive decision support platform that empowers social institutions to design fair and accurate targeting rules tailored to their needs [2].

References

[1] *The State of Social Safety Nets*. The World Bank, 2018.

[2] ANONYMOUS, A. Algorithmic targeting of social policies: Fairness, accuracy, and distributed governance. 2019.

[3] BANK, W. World databank. *World Development Indicators* (2015).

[4] CARBONNEAU, M.-A., CHEPLYGINA, V., GRANGER, E., AND GAGNON, G. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition 77* (2018), 329–353.

[5] CENTER, I. Manual for the implementation of usaid poverty assessment tools. *povertytools. org/training_documents/Manuals/USAID_PAT_Manual_Eng. pdf, accessed 1* (2009).

[6] FISZBEIN, A., AND SCHADY, N. R. *Conditional cash transfers: reducing present and future poverty*. The World Bank, 2009.

[7] HANNA, R., AND OLKEN, B. A. Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries. *Journal of Economic Perspectives 32*, 4 (2018), 201–26.

[8] IBARRARÁN, P., MEDELLÍN, N., REGALIA, F., STAMPINI, M., PARODI, S., TEJERINA, L., CUEVA, P., AND VÁSQUEZ, M. How conditional cash transfers work.

[9] KOENKER, R., AND HALLOCK, K. F. Quantile regression. *Journal of economic perspectives 15*, 4 (2001), 143–156.

A Supplementary materials

Table 2: Relative decrease in exclusion errors obtained by the ML-based versus the status quo methods.

	Colombia											
Region	1	2	3	4	5	6	7	8	9	10	11	12
Reduction %	26.9	24.7	18.2	12.7	23.6	26.7	9.0	12.3	13.4	12.9	24.1	3.1
Region	13	14	15	16	17	18	19	20	21	22	23	24
Reduction %	26.1	11.7	12.5	13.3	16.8	9.5	21.8	25.1	29.7	12.5	28.3	18.2
				Costa Rica								
Region				1	2	3	4	5	6			
Reduction %				25.4	25.1	29.3	20.2	34.1	23.9			