Problem set 2: Predicting income

Big Data and Machine Learning
April 2025

Carlos Manjarres, Juan Felipe Triana, and Maya Gutiérrez

Github link: https://github.com/MayaGutiBan/BDML_PS2_g12

1 Introduction

Poverty remains one of the most pressing global development challenges. In recent decades, substantial progress has been made in reducing poverty rates globally, yet hundreds of millions of people continue to live in conditions of extreme deprivation. Despite improvements in living standards and economic growth, poverty and inequality remain persistent and complex issues, particularly among vulnerable and rural populations (CEPAL, Comisión Económica para América Latina y el Caribe, 2024).

Accurately measuring poverty is central to designing and implementing effective poverty alleviation strategies. However, traditional poverty measurement methods primarily based on detailed household consumption surveys are resource-intensive, time-consuming, and often limited in scope and frequency (Bank, 2018). These limitations constrain governments and organizations in their efforts to rapidly evaluate the impact of social programs and adapt their policies in a timely manner.

The study and measurement of poverty have long been central to the field of development economics. Early economic models often treated poverty as a peripheral concern, focusing instead on aggregate growth and national income. However, a paradigm shift occurred as economists and policymakers increasingly recognized that economic growth alone was insufficient to ensure improvements in individual well-being (Streeten et al., 1981). Central to this transformation was the work of Amartya Sen, who reframed poverty not merely as a lack of income but as a deprivation of capabilities, the real freedoms individuals have to lead the kinds of lives they value (Sen, 1999). These conceptual advances spurred methodological innovations. In particular, the move from unidimensional measures, such as income thresholds, to multidimensional indices reflected an increased appreciation for the complexity of poverty. It laid the groundwork for more nuanced poverty measurement tools, such as the Human Development Index – HDI ((UNDP), 1990) and the Multidimensional Poverty xc ‘ex – MPI (Alkire & Foster, 2011). Household surveys became a crucial tool in capturing this complexity, incorporating questions not only on income and consumption but also on access to services, housing conditions, and educational attainment (Alkire & Foster, 2011). Despite these innovations, the persistent challenges of survey costs, data gaps, and time lags have motivated recent efforts to complement traditional methods with predictive models using big data and machine learning techniques.

Colombia has made important efforts to improve the measurement and monitoring of poverty over the past two decades (Departamento Nacional de Planeación - DNP, 2011). Traditionally, poverty in the country has been assessed using both direct and indirect approaches. The indirect method based on income or consumption thresholds has served as the primary tool for measuring monetary poverty, while complementary tools such as the Unsatisfied Basic Needs Index (NBI) and the Multidimensional Poverty Index (MPI) have captured broader dimensions of deprivation. In line with Amartya Sen’s conceptual distinction between actual deprivations and the capacity to satisfy

basic needs, Colombia's poverty measurement approach has gradually evolved to combine technical rigor with policy relevance.

A major milestone in this evolution was the creation of the Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad – MESEP (Departamento Nacional de Planeación – DNP & Departamento Administrativo Nacional de Estadística – DANE, 2012). This institutional initiative was led by the National Administrative Department of Statistics (DANE) and the National Planning Department (DNP) in response to a break in data comparability caused by the transition from the Continuous Household Survey (ECH) to the Integrated Household Survey (GEIH) in 2006. MESEP addressed these issues by harmonizing the historical series and developing a revised methodology for poverty measurement. Among its contributions were the definition of updated poverty and extreme poverty lines, improvements in the measurement of household income, and the adoption of more recent consumption patterns.

This study uses microdata from the 2018 Monetary Poverty and Inequality Measurement dataset (Departamento Administrativo Nacional de Estadística – DANE, 2018), which is based on the GEIH platform and fully aligned with MESEP's methodological updates. The dataset includes detailed household-level information on income, labor market activity, education, demographic composition, and access to services, making it particularly well-suited for predictive modeling.

Recent literature highlights the growing use of both traditional and non-traditional data sources to improve the accuracy of poverty prediction. Comparisons of classical econometric models and machine learning algorithms using household data show random forest models often deliver more consistent results, but model performance depends on data quality and context (Verme, 2024). Education emerges as a particularly influential factor in predicting poverty, like this study in Costa Rica, echoing similar findings in studies from South Africa and Indonesia, where education level, employment status, and household characteristics such as size and structure significantly influence poverty outcomes (Biyase & Zwane, 2018). Demographic variables like gender and marital status are also essential, with women of reproductive age more likely to live in poverty (Boudet et al., 2018).

Beyond surveys, recent innovations in data science leverage alternative data sources such as satellite imagery, e-commerce behavior, and mobile phone metadata to estimate poverty. For instance, night-time light data has proven effective in detecting poverty hotspots in China, and mobile phone usage patterns help identify ultra-poor households in Afghanistan as accurately as traditional surveys (Verme, 2024); IEEE, 2022). Similarly, the ownership of vehicles such as cars and motorcycles is highly predictive of economic status in Indonesia. These studies emphasize the importance of integrating multiple, often unconventional, data types to capture poverty's complex and dynamic nature. Together, these findings underline that poverty prediction benefits from a combination of socioeconomic, geographic, and behavioral variables, with education, household size, and access to durable goods among the most robust predictors.

The aim of this work is to contribute on poverty prediction by exploring the feasibility of using machine learning techniques to classify households according to their poverty status in Colombia. Leveraging the rich set of socioeconomic and demographic variables captured in the MESEP, we aim to develop predictive models that identify the key features associated with monetary poverty. Our analysis evaluates the performance of several classification algorithms, including Elastic net, class rebalancing and Tree-based models, with the goal of identifying accurate, transparent, and interpretable models that can complement traditional poverty measurement tools.

The results show that while Elastic Net models offered strong interpretability and solid perfor-

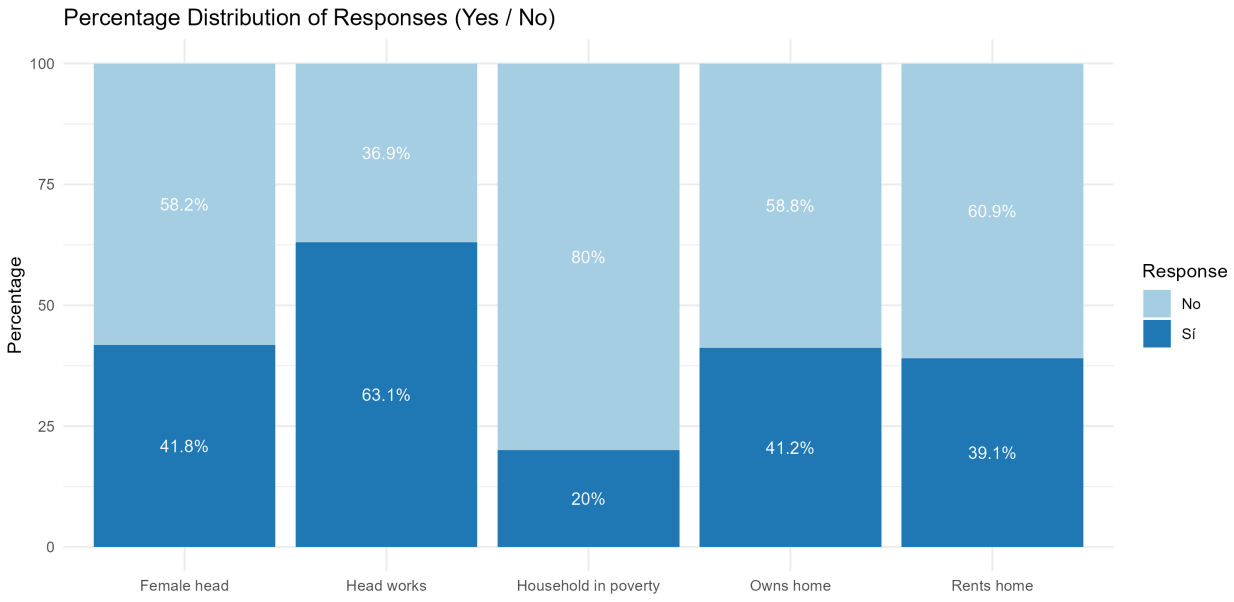

Figure 1: Percentage Distribution of dummy variables

mance, tree-based ensemble methods ultimately delivered superior results, with Random Forests emerging as the best-performing model. These findings underscore the importance of metric-driven model optimization, threshold calibration, and class rebalancing techniques. Overall, the study demonstrates the value of combining rich microdata with carefully selected algorithms and evaluation strategies to improve poverty targeting tools in Colombia.

2 Data description

The Monetary Poverty and Inequality Measurement dataset, offers a solid foundation for predicting household-level poverty in Colombia. The dataset is organized in two complementary components. A household-level file and a person-level file. The household dataset contains essential information on total and per capita household income, expenditure units, and the official poverty and extreme poverty lines used to determine poverty incidence for 2018. Critically, it includes a constructed binary variable, which identifies whether a household falls below the national poverty threshold. This variable serves as the target for our classification task, allowing us to frame the problem as a supervised learning exercise.

The person-level dataset, derived from the 2018 edition of the GEIH, contains individual-level socioeconomic and demographic characteristics necessary for estimating and imputing household income. It includes both original and derived variables used in the official poverty calculation, such as age, education, employment status, and income components. These person-level attributes are essential for constructing additional features and capturing intra-household heterogeneity, thereby enriching the feature space of our models.

Taken together, the structure and content of these datasets make them highly suitable for predictive modeling of poverty. The availability of a clearly defined poverty label, along with a comprehensive set of explanatory variables spanning income, demographics, education, and labor characteristics, provides the necessary conditions for training and evaluating machine learning classifiers. Moreover,

the data's national representativeness and methodological consistency ensure that findings from our models can inform broader policy discussions about poverty targeting and social policy design.

For this process, the dataset includes both training and testing samples at the household and individual levels. The separation between training and testing sets allows us to evaluate the model's performance on an out-of-sample dataset, providing a more reliable measure of its generalization capabilities. It's important to note that for the purpose of the exercise, the testing data was intentionally designed with some missing variables to increase the complexity of the task.

To construct consistent and informative data at the household level, the processing began with the individual-level (person) data. The idea was to derive household-level features by aggregating relevant information across all members of each household. But before doing that, a function was created to generate a standardized set of indicators for each person, such as gender, age group, education, employment status, whether the person is a student or a pensioner, and whether they receive income from various sources like rent, dividends, or subsidies.

Once these indicators were created for each individual, they were aggregated to the household level by summarizing the composition of each household. For example, counting how many members are women, minors, students, workers, or have additional income. Additionally, the educational attainment of the household was summarized using metrics like average years of education and the maximum education level attained by any member.

The data for the head of the household was treated separately to retain their individual characteristics (like gender, education level, employment status) and merge these with the household-level aggregates. This was done for both the training and testing datasets. Finally, a similar preprocessing step was applied to the household-level data. New variables were created to capture aspects of housing conditions and tenure (e.g., whether the house is owned, rented, or held through other arrangements), as well as indicators such as the number of people per room, estimated rent per person, and whether the household pays rent or mortgage installments. This final step ensured that the household-level dataset contained both structural housing information and enriched demographic and socioeconomic characteristics derived from the person-level data.

To gain a first understanding of the population in the dataset, we begin by exploring a set of key dummy variables that capture structural characteristics of the household and housing conditions. These variables, while binary in nature, offer rich insights into the socioeconomic context of the units observed and provide a strong starting point for identifying relevant heterogeneity in the population.

The data allows to identify whether the head of household is female. Approximately 41.8% of households are headed by women. This relatively high share underscores the importance of considering gender-related dynamics in the analysis. Female-headed households often face different labor market opportunities, caregiving responsibilities, and access to social programs, factors that may influence their behavior and socioeconomic outcomes, making this variable potentially relevant for prediction and policy targeting.

Complementing this, we can identify whether the head of the household is currently working. Here, 63.1% of household heads report being employed, suggesting that a significant portion of households rely directly on the labor income of the head. The remaining 36.9%, who do not work, may depend on other household members, informal income sources, or social transfers.

Also, we can determine whether the household rents or owns its home, respectively. Interestingly, 39.1% of households rent their home while 41.2% report owning it. These two groups represent overlapping but not exhaustive categories, and the remaining share may fall under other forms of

tenure such as usufruct, shared housing, or informal arrangements. Housing tenure is often a proxy for wealth accumulation, stability, and exposure to shocks, which may influence household choices and constraints.

According to the household income per capita and the poverty lines in Colombia, it is possible to identify whether the household is officially categorized as being in poverty. 20% of the sample falls under this category. This prevalence is not only critical from a policy perspective but may also interact meaningfully with other household characteristics, such as education, employment, or access to housing. Together, these dummy variables not only reflect important dimensions of socioeconomic status but also offer interpretable features that could support model training and enhance out-of-sample predictive accuracy. Their inclusion is justified not merely for statistical reasons but because they tap into well-documented channels of vulnerability, agency, and constraint within households.

Variable	**Mean**	**Std. Dev.**	**Median**	**Min**	**Max**
Health insurance affiliates	2.53	1.42	2.00	0.00	19.00
Looking for a job	0.08	0.30	0.00	0.00	5.00
Unemployed	0.18	0.45	0.00	0.00	6.00
Students	0.44	0.72	0.00	0.00	9.00
Inactive	1.03	1.03	1.00	0.00	11.00
Aged 60 and over	0.47	0.72	0.00	0.00	6.00
Minors	0.33	0.63	0.00	0.00	9.00
Women	1.74	1.18	2.00	0.00	14.00
Employed	1.50	1.03	1.00	0.00	14.00
Working household members	1.26	0.94	1.00	0.00	9.00
Household size	3.28	1.77	3.00	1.00	28.00
People per room	1.06	0.68	1.00	0.03	13.00

Table 1: Summary statistics of selected household variables

Additionally, the dataset includes a rich set of count variables that detail the number of individuals within each household falling into specific categories. These variables are critical for understanding both the internal structure of the household and its economic capacity, dependency ratios, and potential vulnerability. The average household in the dataset has approximately 3.28 members, with a median of 3 and a maximum of 28 individuals. This wide range suggests notable heterogeneity in household structures, from small nuclear families to extended households or multifamily arrangements. To complement this, the indicator of the number of people per room averages 1.06, with a maximum of 13. This metric is a proxy for overcrowding, a key indicator of housing quality and potential stress. Households report an average of 1.50 employed individuals. This number reflects the primary sources of household income and is closely linked to welfare status. Notably, the standard deviation is relatively high (1.03), indicating significant variation in labor market engagement across households. The average number of unemployed individuals is 0.18, and only 0.08 people per household on average are actively looking for a job. These low figures suggest that outright job-seeking is relatively uncommon in the sample, possibly due to discouraged worker effects or structural labor market exclusion, especially among women or older adults. On average, households have 0.44 students, 0.47 seniors over 60 and 0.33 children. This composition offers a nuanced view of the household life cycle stage and dependency burden. The count of inactive members averages 1.03 per household. This group includes individuals not actively participating in the labor market, such as caregivers, retirees, or discouraged workers. Their presence may affect

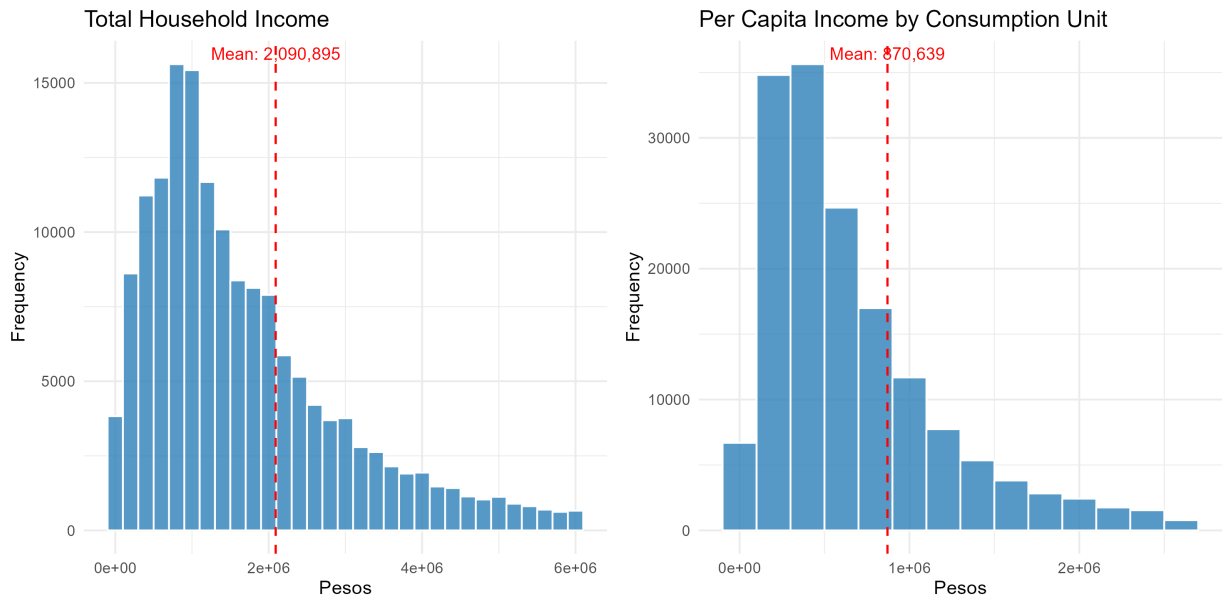

Figure 2: Income distribution by household

household needs (e.g., for subsidies or services) and income-generating potential, which can be critical for targeting social interventions. Finally, the number of women per household averages 1.74, with a range from 0 to 14. Income is a central variable in this analysis, reflecting the household's capacity to meet needs, save, or invest in education and health. Two key indicators are examined, total household income and per capita income adjusted by consumption units. To ensure clearer interpretation and reduce the influence of extreme values, both distributions exclude the top 5% of the income range.

As shown in the left panel of the figure, total household income exhibits a right-skewed distribution, where most households cluster around lower income levels, and a long tail extends toward higher incomes. The mean is just above $2 million Colombian pesos. This skewness is characteristic of income data in developing economies and reflects structural inequalities in the labor market, access to capital, and education. Despite the truncation at the 95th percentile, the spread remains considerable, indicating substantial variation even within the "typical" range of the population. This reinforces the need to account for income heterogeneity in modeling efforts, as it may influence access to welfare programs, vulnerability to shocks, and labor supply decisions. The right panel displays the distribution of per capita income, adjusted for household composition using a consumption unit equivalence scale. This adjustment improves comparability across households of different sizes and structures, as it accounts for economies of scale in consumption. The resulting distribution is even more concentrated around lower values than total household income, with most observations falling below $1 million pesos per person, and a mean of 0.87 million pesos. The sharp left peak reflects the presence of many low-income individuals, while the right tail again shows considerable inequality. This transformation not only helps normalize income but is essential for targeting social programs and poverty identification, which often use per capita metrics. In sum, the richness and diversity of variables available in the dataset, from binary indicators of household structure and housing conditions, to detailed counts of household composition and labor participation, to continuous measures of income; provide a comprehensive view of Colombian households. This multidimensional information allows for nuanced characterizations that reflect both structural

conditions and dynamic economic capacities. Moreover, the presence of well-defined, interpretable features creates fertile ground for predictive exercises aimed at identifying vulnerable populations, forecasting policy impacts, or modeling household behavior. The dataset thus offers a robust foundation for both descriptive analysis and data-driven policy design.

3 Models and Results

Elastic Net model selection

Elastic Net (EN) was used as a model fit containing all p predictors but using a technique that constrains or regularizes the coefficient estimates, or equivalently, that shrinks the coefficient estimates towards zero. This constraint improves the fit by shrinking the coefficient estimates and therefore significantly reduces their variance. EN simultaneously does automatic variable selection and continuous shrinkage, and it can select groups of correlated variables. It is like a stretchable fishing net that retains 'all the big fish'.

Simulation studies and real data examples show that the elastic net often outperforms the lasso in terms of prediction accuracy, because the strict convexity part of the penalty (ridge) solves the grouping instability problem of instability in variable selection with correlated predictors.

Model	Accuracy	Precision	Recall	F1 Score
1. EN_Acu	0.8243	0.7333	0.4728	0.5749
2. EN_F	0.8272	0.7172	0.5160	0.6002
3. EN_F_DCutoff	0.7809	0.5447	0.7835	0.6426
4. EN_F_DCutoff_PROC	0.8043	0.5906	0.7221	0.6497

Table 2: Elastic Net and Tree-Based Model Performance Metrics

In EN, we have two hyperparameters to tune: the mixing parameter between Lasso (=1) and Ridge (=0) and , the penalty strength. The alpha parameter, which determines the balance between L1 (lasso) and L2 (ridge) penalties, To tune the hyperparameters of the elastic net model, we performed a grid search over a predefined range of values for the mixing parameter α and the regularization parameter λ. The parameter α was tested over the range `seq(0, 1, by = 0.2)`. This range includes values from pure ridge regression ($\alpha = 0$) to pure lasso regression ($\alpha = 1$), as well as intermediate combinations.The parameter λ, which controls the strength of the regularization, was tested over the range $10^{\texttt{seq}(10,-2,\text{length}=10)}$. This exponential scale covers a wide span from very strong regularization ($\lambda = 10^{10}$) to very weak regularization ($\lambda = 10^{-2}$), ensuring that the model evaluates both underfitting and overfitting scenarios.These hyperparameters were chosen through bidimensional cross validation. By choosing a grid of values from 0 to 1 and, for each , perform cross-validation over using cv.glmnet().Then, the best , pair based on cross-validation error is selected.

Additionally, model adjusting (model tuning) was performed. Optimizing the model or the model's hyperparameters to improve its ability to detect the minority class. In the case of logistic elastic net regression, this included adjusting the explanatory variables. Changing the target metric, the group attempted to prioritize alternative metrics, such as recall or sensitivity, so that the model maximizes the rate of true positives (poor households classified correctly). By defining a custom evaluation function, which includes key metrics such as sensitivity, precision, and the F1 metric, we were able to compare statistics and decide on the optimal tuning parameter. This allowed us to optimize the hyperparameters of the regularized logistic regression model, looking for the best

Lasso/Ridge penalty combination to maximize the chosen metric.

The metric that was chosen got the majority of models, given that it was the metric used to score predictions in the Kaggle competition, was the F1-score. The trade-off between precision and recall is measured with the F1-score, which is the harmonic mean of both metrics. The objective is to find the threshold that maximizes the F1-score, as this gives us the best balance between capturing the minority class (recall) without losing too much precision.

Another modification that increased the performance of the model was the choice of classification threshold (alternative cutoffs). The standard threshold of 0.5 may be inadequate, especially for unbalanced data. Therefore, in some cases, we reduced the threshold (e.g. 0.3) to classify more cases as poor, increasing sensitivity. Additionally, we adjust the threshold based on metrics such as the ROC curve, balancing precision and sensitivity, which improved the F1 score from 0.60 to 0.64.

By modifying the classification threshold, we change the definition of a positive event (poor), adjusting the rate of false positives (FP) and true positives (TP) of the model. To find the optimal threshold, we will use the ROC (Receiver Operating Characteristic) curve and select a cut-off point that achieves a balance between detecting as many poor households as possible (TP) without excessively increasing false alarms (FP). A lower threshold increases true positives (TP), reducing the risk of classifying poor households as not poor. A lower threshold can also increase false positives (FP), i.e. classifying some non poor households as poor. Maximizing the AUC area in the ROC curve provides a balance where we improve the detection of poor households without generating too many false positives.

In classification problems with unbalanced classes, the ROC Curve may not be the best metric for evaluating the model, as it does not directly consider performance in the minority class. Instead, we can use the Precision-Recall Curve, which evaluates the balance between the precision and recall (TPR, sensitivity) of the model.

Class Rebalancing:

Modifing the composition of the training set to make the model learn with a more balanced dataset also led to an increase in the F1 score (to 0.6497 in the case of the Elastic Net Logistic Regression models estimated). There where several ways of rebalancing the quantity of observations in the poor and not poor clases, for example re-weighting observations (assigning greater weight to the minority class. This forces the model to pay more attention to cases of unemployment and reduces the bias towards the majority class), subsampling (reducing the number of observations of the majority class (poor), to prevent them from dominating the learning), and over-sampling (artificially increase the number of observations from the minority class (e.g. by replicating poor household cases). However, we chose one advanced techniques, SMOTE (Synthetic Minority Over-sampling Technique), which generates synthetic examples in the minority class instead of simply replicating them. This led to the highest F1 score given that this algorithm helps to overcome the overfitting problem posed by random oversampling. It focuses on the feature space to generate new instances with the help of interpolation between the positive instances that lie together, by matching the observation to its K nearest neighbors.

Tree-based models:

The best performing model in this analysis was a Random Forest (RF) classifier, tuned using 5-fold cross-validation, with an optimized threshold. The training process involved using the ranger package with a grid search over key hyperparameters to optimize the model's performance, specifically targeting the F1-score to balance precision and recall in the presence of class imbalance.

Model	Accuracy	Precision	Recall	F1 Score
5. TREES_AUC	0.6002	0.6426	0.6497	0.5203
6. XGB GRID PROC	0.8672	0.6471	0.7400	0.6904
7. Bagging	0.8821	0.7356	0.6413	0.6852
8. RF GRID DCutoff PROC	0.8678	0.6401	0.7756	0.7014
9. RF Opti GRID	0.8778	0.7623	0.5663	0.6499
10. Boosting XGB	0.8766	0.7450	0.5834	0.6544

Table 3: Performance metrics of various tree models

The optimized hyperparameters were the following:

Number of random variables considered at each split (mtry): Tuning this hyperparameter involves a classic bias-variance trade-off. A higher (mtry) allows each tree to consider more variables when splitting, which can result in lower bias but also higher variance, as the trees may become more similar and overfit the training data. Conversely, a lower (mtry)increases tree diversity (lower correlation among trees), thereby reducing variance but potentially increasing bias due to weaker individual trees. In our tuning process, we tested (mtry) values of 2, 4, 6, and 8. The optimal value was 4, which achieved the best F1-score during cross-validation.

Minimum number of observations required in terminal nodes (min.node.size): This hyperparameter also reflects a bias-variance trade-off. A larger (min.node.size) results in simpler, shallower treeswith higher bias but lower variance, while a smaller (min.node.size) produces more complex trees that capture more data patterns (lower bias), but at the risk of overfitting (higher variance). We evaluated values of 3, 5, and 8, and found that the optimal setting was 5, striking a good balance and producing the best F1-score.

Threshold Optimization via ROC PROC Curve: Rather than relying on the default 0.5 threshold, a custom classification cutoff was selected by analyzing the ROC curve and computing the precision-recall tradeoff (PROC). A sequence of thresholds was evaluated, and the threshold that maximized the F1-score was selected. This decision rule significantly improved the classifier's ability to correctly identify the minority class (i.e., those labeled as "Pobre").

As shown in the results table, applying this optimized threshold significantly improved the model's performance in identifying the minority class, enhancing both precision and recall. This methodological choice helped mitigate class imbalance without requiring sampling techniques or class weighting, focusing instead on threshold calibration and F1-score optimizatio. Also the table display that random forest model with, optimized threshold its the best performing model.

4 Conclusions

This study highlights the potential of machine learning techniques to enhance poverty measurement efforts by leveraging rich, multidimensional household-level data. While traditional survey-based approaches remain vital for capturing detailed information on household welfare, they are often costly, infrequent, and slow to adapt. In contrast, predictive models such as Random Forest or Elastic Net offer a scalable and interpretable way to identify poor households with considerable accuracy, particularly when trained on nationally representative datasets like the 2018 GEIH and aligned with MESEP's methodological framework.

While Elastic Net models performed well, tree-based ensemble methods ultimately outperformed

them in terms of F1-score. Random Forests, in particular, emerged as the top-performing model, especially after hyperparameter tuning and threshold optimization based on ROC analysis. This model achieved a strong balance between detecting poor households (recall) and maintaining classification precision. Overall, the findings highlight the importance of metric-driven model optimization, classification threshold calibration, and class rebalancing when addressing imbalance.

Our results confirm that well-engineered features derived from individual and household-level socioeconomic data, especially those capturing education, income, labor status, and household composition can be powerful predictors of monetary poverty. Moreover, optimizing classification thresholds and tuning model hyperparameters significantly improves performance, especially in unbalanced datasets where poor households represent a minority class. The F1-score, as a metric balancing precision and recall, proved particularly useful in guiding model selection and evaluation.

Expanding this framework to estimate poverty dynamics over time or to incorporate non-monetary dimensions of poverty would align the modeling approach more closely with capability-based perspectives on human development. Ultimately, by combining rigorous statistical methods with a strong understanding of the socioeconomic context, predictive modeling can become a powerful ally in the design of responsive and inclusive public policies aimed at reducing poverty.

References

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, *95*(7–8), 476–487. https://doi.org/10.1016/j.jpubeco.2010.11.006

Bank, W. (2018). *Poverty and shared prosperity 2018: Piecing together the poverty puzzle*.

Biyase, M., & Zwane, T. (2018). An empirical analysis of the determinants of poverty and household welfare in south africa. *The Journal of Developing Areas*, *52*(1), 115–130. https://doi.org/10.1353/JDA.2018.0008

Boudet, A. M. M., et al. (2018). *Gender differences in poverty and household composition through the life-cycle: A global perspective* (tech. rep.). World Bank. Washington, DC. https://doi.org/10.1596/1813-9450-8360

CEPAL, Comisión Económica para América Latina y el Caribe. (2024). *Panorama social de américa latina y el caribe, 2024: Desafíos de la protección social no contributiva para avanzar hacia el desarrollo social inclusivo* [LC/PUB.2024/21-P, 267 páginas]. CEPAL.

Departamento Administrativo Nacional de Estadística – DANE. (2018). *Medición de pobreza monetaria y desigualdad 2018* (tech. rep.).

Departamento Nacional de Planeación - DNP. (2011). *Índice de pobreza multidimensional (ipm-colombia) 1997-2008 y meta del pnd para 2014* (tech. rep.).

Departamento Nacional de Planeación – DNP & Departamento Administrativo Nacional de Estadística – DANE. (2012). *Misión para el empalme de las series de empleo, pobreza y desigualdad (mesep)* (tech. rep.).

Sen, A. (1999). *Development as freedom*. Knopf.

Streeten, P., Burki, S. J., ul Haq, M., Hicks, N., & Stewart, F. (1981). *First things first: Meeting basic human needs in developing countries*. Oxford University Press.

(UNDP), U. N. D. P. (1990). *Human development report 1990: Concept and measurement of human development*. Oxford University Press.

Verme, P. (2024). Predicting poverty [lhae044]. *The World Bank Economic Review*. https://doi.org/10.1093/wber/lhae044