ISSN: 2658-8455
Volume 6, Issue 9 (2025), pp. 303-318.


# Targeting Social Assistance Beneficiaries Using Machine Learning: A Poverty Probability-Based Approach

# Ciblage des bénéficiaires de l'aide sociale par l'apprentissage automatique: Une approche fondée sur la probabilité de pauvreté

**Chaymae SAHRAOUI**, (***PhD candidate)***
*Research Laboratory in Applied Mathematics in Economic and Management*
*Faculty of Law, Economics and Social Sciences of Ain Sebaa*
*UniversityHassan II of Casablanca, Morocco*

**Tarek ZARI**, (***Full Professor)***
*Research Laboratory in Applied Mathematics in Economic and Management*
*Faculty of Law, Economics and Social Sciences of Ain Sebaa*
*UniversityHassan II of Casablanca, Morocco*

| | |
|---|---|
| **Correspondence address :** | Faculty of Law, Economics and Social Sciences of Ain Sebaa<br>BP : 2634، Route des Chaux et Ciments Beausite, Casablanca 20254<br>+2125223-43482 |
| **Disclosure Statement :** | The authors declare that they have not received any financial support that could have influenced the objectivity of this study. They take full responsibility for any potential plagiarism and for the accuracy of the results presented in this article. |
| **Conflict of Interest :** | The authors report no conflicts of interest. |
| **Cite this article :** | SAHRAOUI, C., & ZARI, T. (2025). Targeting Social Assistance Beneficiaries Using Machine Learning: A Poverty Probability-Based Approach. *International Journal of Accounting, Finance, Auditing, Management and Economics*, *6*(9), 303–318. |
| **License** | **This is an open access article under the CC BY-NC-ND license** |

*Received: 20/06/2025* *Accepted: 04/08/2025*

# Targeting Social Assistance Beneficiaries Using Machine Learning: A Poverty Probability-Based Approach

## Abstract :

In a context where social inequalities are deepening and public resources are becoming increasingly scarce; the fair and effective identification of social assistance beneficiaries has become a central issue. Traditional targeting methods, such as categorical eligibility or proxy means testing, are now showing their limits, frequently producing inclusion and exclusion errors.

This study relies on a synthetic dataset of 12,600 individuals described by 59 socio-economic variables, ranging from demographic characteristics and education level access to financial and digital services. Three supervised learning models are compared: logistic regression, Random Forest, and XGBoost. The results reveal that tree-based models outperform logistic regression, particularly in reducing exclusion errors, which are especially critical in social policy contexts.

The analysis of key variables highlights the decisive role of education levels, place of residence (urban/rural), and access to digital and financial services. These findings confirm the need for a multidimensional approach to poverty that goes beyond purely monetary criteria. Finally, the study emphasizes the ethical challenges raised using algorithms: transparency, bias reduction, and institutional accountability emerge as essential conditions for legitimizing their integration into social protection and for contributing to more inclusive and equitable systems.

**Keywords:** Algorithmic targeting; Social protection; Machine learning; Multidimensional poverty; Data ethics.
**Classification JEL:** I32; I38; C45; C55; H53
**Paper type:** Empirical Research

## Résumé :

Dans un contexte où les inégalités sociales s'aggravent et où les ressources publiques se raréfient, la question de l'identification juste et efficace des bénéficiaires de l'aide sociale devient centrale. Les méthodes de ciblage classiques, comme l'éligibilité catégorielle ou le proxy means testing, montrent aujourd'hui leurs limites, en produisant fréquemment des erreurs d'inclusion ou d'exclusion.

Cette étude mobilise un jeu de données synthétique de 12 600 individus décrits par 59 variables socio-économiques, allant des caractéristiques démographiques au niveau d'instruction, en passant par l'accès aux services financiers et numériques. Trois modèles d'apprentissage supervisé ont été comparés : la régression logistique, la forêt aléatoire (Random Forest) et XGBoost. Les résultats révèlent que les modèles fondés sur les arbres offrent de meilleures performances, notamment pour réduire les erreurs d'exclusion, particulièrement sensibles dans les politiques sociales.

L'analyse des variables déterminantes met en évidence le rôle décisif du niveau d'éducation, du lieu de résidence (urbain/rural) et de l'accès aux services numériques et financiers. Ces constats confirment l'importance d'une approche multidimensionnelle de la pauvreté, dépassant le seul critère monétaire. Enfin, l'étude souligne les enjeux éthiques liés à l'usage des algorithmes : transparence, réduction des biais et responsabilité institutionnelle apparaissent comme des conditions indispensables pour légitimer leur intégration dans la protection sociale et contribuer à des systèmes plus inclusifs et équitables.

.

**Mots clés : Ciblage** algorithmique, Protection sociale, Apprentissage automatique, Pauvreté multidimensionnelle, Éthique des données.
**JEL Classification:** I32; I38; C45; C55; H53
**Type d'article :** Recherche empirique

ISSN: 2658-8455
Volume 6, Issue 9 (2025), pp. 303-318.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

## 1. Introduction

Social protection systems around the world are increasingly confronted with complex challenges. Growing economic volatility, labor market informality, demographic shifts, and the intensification of social inequalities have heightened the urgency of designing systems that are both inclusive and efficient. In many low- and middle-income countries (LMICs), a large share of the population remains either unprotected or poorly targeted by existing programs, resulting in significant exclusion errors—where the poor are left out—and inclusion errors—where non-poor individuals benefit unduly (Brown & Ravallion, 2020). Moreover, crises such as the COVID-19 pandemic have exposed the fragility of traditional targeting mechanisms and intensified demands for greater responsiveness and adaptability in social protection (Gentilini et al., 2022).

In this context, improving the precision and responsiveness of social assistance targeting mechanisms has become a central concern for policymakers, development partners, and data scientists alike. The aim of this study is to empirically assess the potential of machine learning techniques to enhance the targeting accuracy of social assistance programs. By comparing the predictive performance of logistic regression, Random Forest, and XGBoost on a synthetic dataset reflecting realistic socio-economic profiles, the study seeks to determine which algorithm offers the best balance between sensitivity and specificity. Additionally, it analyzes the key features influencing eligibility classification and discusses the ethical and policy implications of deploying such models in real-world welfare systems.

Traditional approaches to targeting, such as categorical eligibility and proxy means testing (PMT), have been widely used due to their simplicity and administrative feasibility. However, their limitations are well documented. Categorical schemes are often too rigid to capture the heterogeneity of poverty, while PMT relies on a limited number of observable household attributes, making it vulnerable to manipulation and misclassification (Hanna & Olken, 2018). Moreover, these approaches often lack the flexibility to adapt to changing social and economic conditions, especially in times of crisis or during rapid transitions in the labor market (Beegle et al., 2018).

In recent years, the growing availability of socio-economic data and advances in artificial intelligence have paved the way for more dynamic and data-driven approaches to welfare targeting. Machine learning (ML) offers the ability to uncover complex patterns and interactions within high-dimensional data, thereby enabling more accurate predictions of welfare status or risk of poverty. Studies such as Athey (2017) and Levy et al. (2021) have highlighted the potential of ML models to outperform traditional statistical techniques in classification tasks, particularly when dealing with noisy or incomplete datasets. These approaches also allow for the continuous updating of predictive models based on new data flows, thus enhancing responsiveness and operational efficiency (Abebe et al., 2020; Athey & Imbens, 2019).

Furthermore, early empirical implementations of ML-based targeting in public policy have shown promising results. The Togo Novissi program leveraged mobile metadata and AI algorithms to identify vulnerable informal workers during the COVID-19 pandemic (Aarvik, 2021). Brazil's Cadastro Único has used machine learning to improve consistency checks and detect anomalies in its social registry (World Bank, 2022). These examples demonstrate that beyond algorithmic accuracy, ML offers operational advantages such as scalability, automation, and the ability to continuously learn from new data.

In research environments where access to real administrative data is restricted, the use of high-quality synthetic datasets has become an accepted practice for simulating welfare scenarios and evaluating model robustness (Chen et al., 2021; El Emam et al., 2020). This study adopts such

an approach, ensuring that the simulated data preserves key statistical properties found in real-world poverty registries.

Despite this potential, the deployment of machine learning in social protection also raises important questions regarding model transparency, fairness, and the interpretability of decision factors that are essential in the design of equitable and accountable social policy systems. Therefore, empirical studies that not only evaluate the predictive performance of ML models, but also examine their interpretability and practical implications, are crucial.

This paper contributes to this evolving field by empirically assessing the predictive capacity of three supervised machine learning models—logistic regression, random forest, and XGBoost—for identifying individuals likely to be eligible for social assistance, based on a synthetic dataset simulating real-world socio-economic characteristics. The study places particular emphasis on model comparison, variable importance analysis, and the operational trade-offs inherent in setting decision thresholds. In doing so, it aims to support the development of more adaptive, transparent, and data-informed social protection systems.

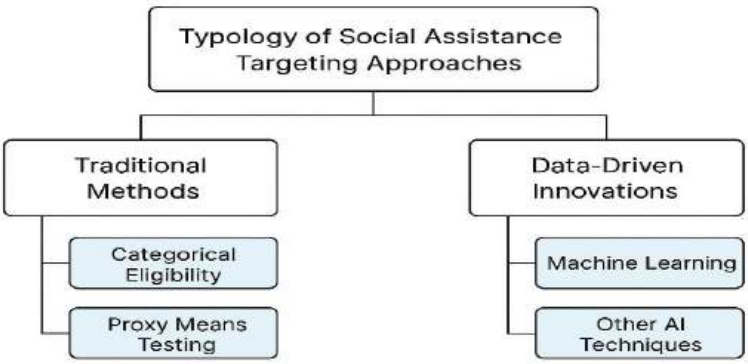
## 2. Literature review and hypothesis development

### 2.1 Literature review

Academic research on targeting mechanisms in social protection has evolved significantly over the past two decades. Early work primarily examined the effectiveness of proxy means testing (PMT) and categorical eligibility schemes, emphasizing their limitations in accounting for the multidimensional and dynamic nature of poverty. Brown and Ravallion (2020) documented widespread inclusion and exclusion errors across developing countries under PMT-based approaches, while Hanna and Olken (2018) compared targeted versus universal transfer schemes, highlighting trade-offs between efficiency and equity. Further critiques have underlined how PMT may entrench social exclusion, especially when household characteristics poorly reflect transient poverty or vulnerability (Kidd & Athias, 2020; Beegle et al., 2018).

A typology of targeting approaches is summarized in Figure 1, distinguishing traditional techniques from more recent data-driven innovations.

***Figure 1. Typology of Social Assistance Targeting Approaches.***




***Source: Based on Beegle et al. (2018), Kidd & Athias (2020), and Aiken et al. (2021).***

With the rapid growth of digital infrastructures and the increasing availability of socio-economic data, researchers have turned their attention to machine learning (ML) as a way to improve the precision of targeting in social protection. Athey (2017) argued that ML techniques often outperform traditional econometric tools, particularly in settings where data are highly dimensional and relationships are nonlinear. Recent empirical evidence supports this view. Aiken et al. (2022), for example, showed that ML models built on mobile phone data substantially enhanced poverty targeting in Afghanistan. Likewise, Kim (2021) found that

ISSN: 2658-8455
Volume 6, Issue 9 (2025), pp. 303-318.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

Random Forest and Gradient Boosted Trees yielded more accurate poverty predictions than logistic regression in Costa Rica. More recently, Salvador (2024) confirmed the effectiveness of boosting algorithms such as XGBoost and CatBoost in classifying household poverty in the Philippines.

Beyond these performance gains, the literature has broadened to address methodological and operational issues. Scholars have emphasized the importance of sound variable selection, ensuring generalizability, and implementing robust validation methods to avoid overfitting and to strengthen real-world applicability (Chouldechova & Roth, 2020). At the same time, concerns about fairness and accountability have become more prominent. Wachter, Mittelstadt, and Floridi (2017) and the OECD (2022) underline the risks of opacity in complex models, while Eubanks (2018) and Binns (2020) highlight the dangers of bias propagation and algorithmic exclusion when ML is applied in welfare systems.

In light of these debates, a multidisciplinary consensus is emerging that stresses the need for a policy-oriented use of ML. Rather than focusing solely on predictive accuracy, scholars argue that ML applications in social protection should be aligned with ethical standards, institutional realities, and social equity objectives (Narayanan & Vallor, 2021; Kroll et al., 2017).

### 2.2 Hypothesis development

Although prior literature highlights the limitations of traditional targeting approaches such as PMT, recent studies show that machine learning can improve predictive performance and reduce misclassification errors. However, evidence also suggests that predictive gains often come at the expense of interpretability and institutional legitimacy (Wachter et al., 2017; Eubanks, 2018). In developing countries, this trade-off is particularly relevant, as social protection reforms face both fiscal constraints and the need to ensure transparency and fairness in beneficiary selection.

Education, residence, and access to financial or digital services emerge consistently in literature as strong predictors of poverty and social vulnerability (Brown & Ravallion, 2020; Aiken et al., 2022; Kim, 2021). Educational attainment is positively associated with socio-economic mobility and thus expected to influence eligibility outcomes. Similarly, rural households tend to face structural disadvantages, which increases their likelihood of being classified as eligible for assistance. Finally, access to financial and digital services has been identified as an important proxy for welfare inclusion, potentially offering stronger predictive power than demographic variables alone (Salvador, 2024). Based on the above arguments, the following hypotheses are proposed:

- **H1.** Higher educational attainment is positively and significantly associated with the probability of being classified as eligible for social assistance.
- **H2.** Rural residence significantly increases the probability of eligibility compared to urban residence.
- **H3.** Access to financial and digital services provides stronger predictive power for poverty targeting than demographic or income-related variables alone.

## 3. Methodology

### 3.1. Dataset and Target Definition

This research uses a synthesized dataset released via an international data science competition in social assistance targeting ( [Kaggle Poverty Probability Challenge]). The database provides anonymized data of the socio-economic profiles of individuals in a number of fictitious countries, with 12,600 instances and 59 variables. The covariates such as age, literacy and

numeracy ability, working status, and use of banking or digital services are simulated based on real-world poverty dynamics.

***Table 1[1] Descriptive statistics for key predictors.***

| ***Variable*** | ***Description*** | ***Type*** | ***Value / %*** |
|---|---|---|---|
| Age | Age of respondent (years) | Continuous | Mean = 36.3, SD = 15.1 (min = 15, max = 115) |
| Residence (is_urban) | 1 = Urban, 0 = Rural | Binary | Rural = 67.1 %, Urban = 32.9 % |
| Gender (female) | 1 = Female, 0 = Male | Binary | Female = 55.8 %, Male = 44.2 % |
| Marital status (married) | 1 = Married, 0 = Not married | Binary | Married = 64.9 %, Not married = 35.1 % |
| Literacy | 1 = Literate, 0 = Illiterate | Binary | Literate = 61.4 %, Illiterate = 38.6 % |
| Employment last year | 1 = Employed, 0 = Not employed | Binary | Employed = 58.9 %, Not employed = 41.1 % |
| Financial inclusion | 1 = Has access to banking/financial services, 0 = otherwise | Binary | Included = 49.3 %, Not included = 50.7 % |
| Digital access (can_use_internet) | 1 = Has access to digital services (phone/internet), 0 = otherwise | Binary | Access = 45.0 %, No access = 55.0 % |
| Education level | 0 = None, 1 = Primary, 2 = Secondary, 3 = Higher | Categorical | None = 20.6 %, Primary = 36.8 %, Secondary = 33.0 %, Higher = 9.6 % |

***Source: Authors***

The target is the "poverty probability" as calculated in the original dataset. Using a binary threshold for classification purposes: who has the probability of ≥ 0.5 is eligible to receive social assistance. Although this threshold is often set to balance sensitivity and specificity, in practice, it may be modified to account for specific policy goals, particularly when minimizing no-shows is the priority (Chen et al., 2021; Grover et al., 2020).

Nevertheless, the dataset has significant drawbacks. As an artificial body, it does not stand for any actual national population, and it should not be used to reflect the socio-economic character of any actual national population. It should not be mistaken as a true administrative data set; instead, it is a reference that allows testing the performance of machine learning and econometric models in a more controlled environment when real administrative data is prevented by privacy or institutional restrictions (El Emam et al., 2020; Goncalves et al., 2020).

### 3.2. Feature Engineering and Preprocessing

The dataset includes both numerical and categorical features. Categorical variables were converted into binary indicators using one-hot encoding—a standard approach to preserve model compatibility and avoid ordinal bias (Zheng & Casari, 2018). All missing values were treated with zero imputation, a strategy commonly used in conjunction with tree-based algorithms, which are inherently robust to unscaled or sparse data (Biau & Scornet, 2016). Feature alignment ensured consistency across training and validation subsets. No normalization was applied, as tree-based models are invariant to feature scaling.

The final dataset was split into training and validation sets using an 80/20 ratio, consistent with standard supervised learning protocols.

### 3.3. Model Selection

Three supervised classification models were selected for comparison:

**Logistic Regression.** Logistic regression is a linear model widely used in binary classification tasks and serves as a baseline due to its interpretability and low variance. Formally, the model

[1] *Notes: Table 1 presents summary statistics of the key measures employed in the analysis. The dataset consists of 59 independent variables; only the most significant predictors are listed here for clarity*

ISSN: 2658-8455
Volume 6, Issue 9 (2025), pp. 303-318.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

is specified as follows:

$$Log\left(\frac{P(Y_i = 1)}{1 - P(Y_i = 1)}\right) = \beta_0 + \beta' X_i + \varepsilon_i$$

Where:

- $Y_i$ is the dependent variable equal to 1 if individual $i$ is classified as poor (poverty probability ≥ 0.5), and 0 otherwise.
- $X_i$ is the vector of explanatory variables, including demographic, educational, employment-related, and access-to-services characteristics.
- $\beta$ is the vector of coefficients associated with these predictors.

- **Random Forest:** a bagging ensemble method known for its robustness and ability to handle nonlinear interactions (Breiman, 2001).
- **XGBoost:** an efficient implementation of gradient boosting algorithms, particularly suited for structured/tabular data and widely adopted in applied machine learning (Chen & Guestrin, 2016).

All models were trained using default hyperparameters, aligning with the goal of establishing baseline comparisons. Future iterations may benefit from hyperparameter tuning via grid search or random search to optimize performance across metrics (Bergstra & Bengio, 2012).

### 3.4. Evaluation Metrics

Model performance was evaluated using standard classification metrics:

- Accuracy: overall correctness of the classification.
- Precision: proportion of true positives among predicted positives.
- Recall: proportion of true positives among actual positives.
- F1-score: harmonic mean of precision and recall, offering a balanced metric in the presence of class imbalance.

We also examined confusion matrices and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) to understand the trade-offs across different classification thresholds. These metrics are particularly relevant in social protection applications, where reducing exclusion errors (false negatives) is often a central policy priority (Levy et al., 2021; Barocas et al., 2019).

## 4. Results and Discussion

This section details the empirical results derived from three supervised machine learning models (logistic regression, random forest, and XGBoost) to classify the eligibility for social support. The investigation looks further into a comparison of model screening performance using accuracy, precision, recall, and F1-score as evaluation measures, followed by the process of feature importance determination to identify predictive variables that account most for classifying into poor-based eligibility.

### 4.1. Model Performance Comparison

Table 2 clearly illustrates the relative strengths of the three models. Random Forest stands out with the highest accuracy (0.778, 95% CI [0.767–0.787]) and the strongest recall for identifying eligible households (0.886, 95% CI [0.869–0.903]), which is essential to avoid leaving out those most in need. XGBoost produces very similar outcomes, trailing slightly in accuracy but achieving better recall for non-eligible households (class 0), thus lowering the risk of granting benefits to ineligible groups.

***Table 2 summarizes the classification performance of each model on the validation dataset.***

| Metric | Logistic Regression | Random Forest | XGBoost |
|---|---|---|---|
| Accuracy | 0.762 [0.751 – 0.773] | 0.778 [0.767 – 0.787] | 0.769 [0.758 – 0.785] |
| Precision (0) | 0.700 [0.677 – 0.725] | 0.738 [0.713 – 0.766] | 0.702 [0.680 – 0.738] |
| Precision (1) | 0.788 [0.780 – 0.797] | 0.793 [0.786 – 0.802] | 0.799 [0.790 – 0.808] |
| Recall (0) | 0.580 [0.561 – 0.600] | 0.582 [0.561 – 0.606] | 0.611 [0.583 – 0.637] |
| Recall (1) | 0.863 [0.847 – 0.878] | 0.886 [0.869 – 0.903] | 0.856 [0.839 – 0.881] |
| F1-score (0) | 0.635 [0.618 – 0.651] | 0.651 [0.634 – 0.667] | 0.653 [0.637 – 0.671] |
| F1-score (1) | 0.824 [0.815 – 0.833] | 0.837 [0.828 – 0.844] | 0.827 [0.818 – 0.840] |
| Macro F1 | 0.729 [0.716 – 0.741] | 0.744 [0.732 – 0.755] | 0.740 [0.728 – 0.756] |
| Weighted F1 | 0.756 [0.745 – 0.767] | 0.771 [0.760 – 0.781] | 0.765 [0.754 – 0.780] |

***Source: Authors***

By contrast, logistic regression performs less well across most indicators. Although its interpretability makes it a valuable benchmark, its recall and F1-scores reveal its limits when dealing with the complex, non-linear relationships that characterize poverty dynamics. Tree-based methods, by design, handle this heterogeneity more effectively.

Taken together, the results suggest that Random Forest offers the best overall balance between accuracy and sensitivity. This conclusion echoes earlier work highlighting the comparative advantage of ensemble models when applied to socio-economic datasets with high variability (Athey & Imbens, 2019; Abebe et al., 2020).

### 4.2. Feature importance analysis

To gain deeper insights into the determinants of eligibility predictions, we analyzed feature importance rankings produced by the three models used in the study: Logistic Regression, Random Forest, and XGBoost. These visualizations highlight which variables contributed most significantly to the classification outcomes. Feature importance is a widely adopted technique in supervised learning to assess the relative contribution of input variables to model decisions (Zheng & Casari, 2018; Molnar, 2022).

**Figure 2** displays the ten most influential features according to the Logistic Regression model, ranked by the absolute values of their standardized coefficients. The leading predictors include country-level indicators such as *country_D* and *country_A*, followed by variables like *religion_N*, *urban residence*, *marital status*, and the relationship to the household head. These results reflect the tendency of linear models to emphasize broader socio-demographic and geographic structures, offering transparency but often failing to capture more complex interactions (Biau & Scornet, 2016).

**Figure 3** presents the top predictors identified by the Random Forest model. This ensemble method highlights features such as *age*, *education level*, *urban status*, and access to digital and financial services (e.g., *phone_technology*, *share_hh_income_provided*). These results suggest that non-linear models can better capture the multidimensional nature of vulnerability, extending beyond income-based measures to reflect behavioral and access-related characteristics (Beegle et al., 2018).

**Figure 4** shows the top-ranked variables according to the XGBoost model. Like Random Forest, this model identifies *education level*, *urban location*, and *access to banking and internet services* as key predictors. It also underscores the relevance of geographic indicators (*country_A*, *country_D*) and behavioral traits such as *can_use_internet* and *active_bank_user*, supporting recent findings on the importance of digital inclusion in welfare eligibility (Beegle, 2018).

Overall, the convergence on certain variables—such as education level and urban status—confirms their centrality in shaping eligibility outcomes. Meanwhile, divergences in the rankings highlight the influence of model architecture on variable importance. Tree-based models like Random Forest and XGBoost are better suited to uncover non-linear patterns and

ISSN: 2658-8455
Volume 6, Issue 9 (2025), pp. 303-318.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

interaction effects without the need for explicit specification (Breiman, 2001; Chen & Guestrin, 2016).

***Comparative Feature Importance across Three Classification Models.***

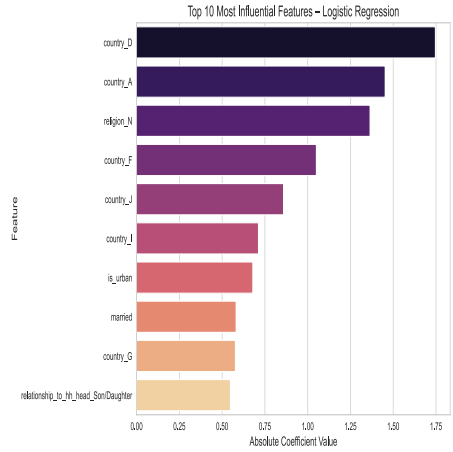



***Figure 2. Top 10 Most Important Features According to the Logistic Regression Model***

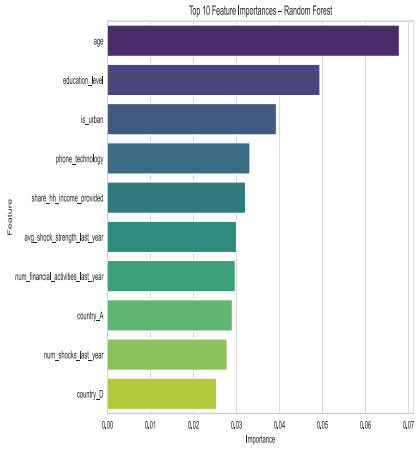



***Figure 3. Top 10 Most Important Features According to the Random Forest Model***

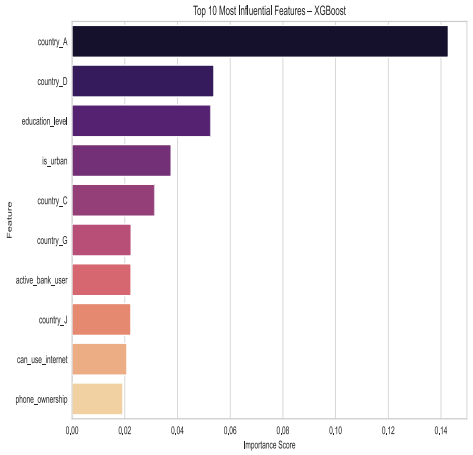
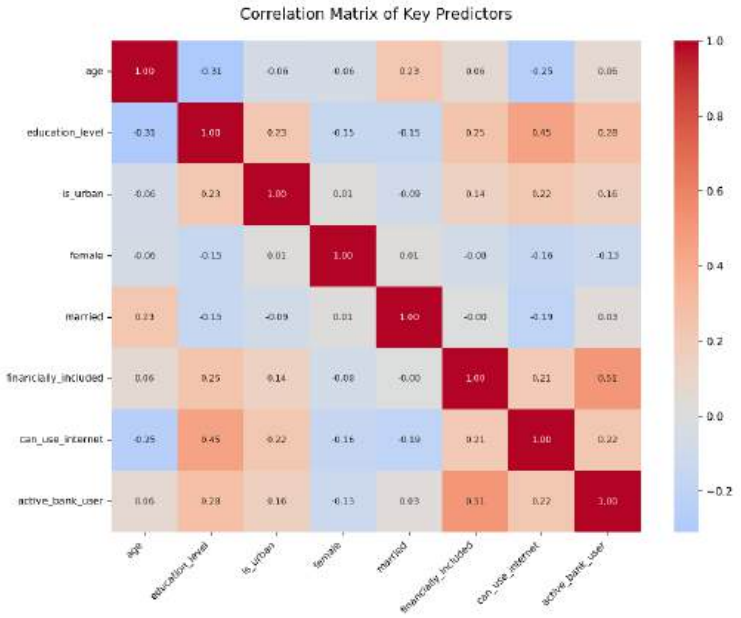



***Figure 4. Top 10 Most Important Features According to the XGBoost Model***

## Correlation between key predictors

To complement the feature importance results, we examined the correlations among the main socio-economic predictors. Figure 5 presents the Pearson correlation matrix for a subset of variables frequently highlighted across the three models (*age*, *education_level*, *is_urban*, *married*, *financially_included*, *can_use_internet*, *active_bank_user*).

***Figure 5. Correlation Matrix of Key Predictors***




***Source: Authors***

The results indicate generally weak to moderate correlations. The strongest association appears between financial inclusion and active bank use ($r \approx 0.51$), reflecting the intuitive link between access to financial services and active usage. A moderate correlation is also observed between education level and digital access ($r \approx 0.45$), suggesting that higher education is associated with

greater use of digital technologies. Negative correlations, such as between age and education level ($r \approx -0.31$), highlight expected generational gaps in education attainment.
Overall, no correlation exceeds 0.6, which indicates that multicollinearity is not a major concern in the models. These results strengthen the robustness of the feature importance analysis, confirming that the selected predictors provide complementary—rather than redundant—information to the classification task.

### 4.3. Confusion Matrix and ROC Curve Analysis

To further assess classification performance, we analyzed and compared the confusion matrices and receiver operating characteristic (ROC) curves across all three models—Logistic Regression, Random Forest, and XGBoost. These tools provide complementary insights into the models' ability to distinguish between eligible (class 1) and non-eligible (class 0) individuals, beyond overall accuracy metrics (Hand & Till, 2001).

***Comparative Confusion Matrix across Three Classification Models.***

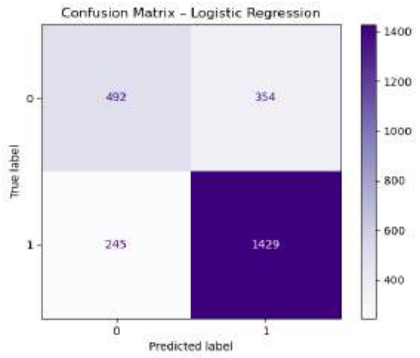



***Figure 6 :Confusion Matrix -Logistic Regression Model***

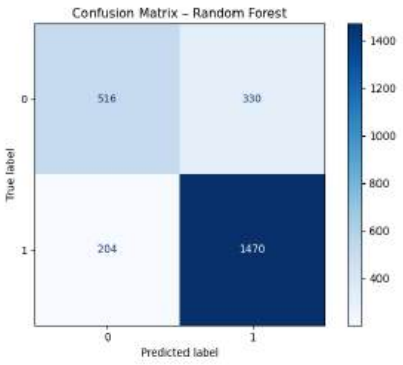



***Figure7 :Confusion Matrix - Random Forest Model***

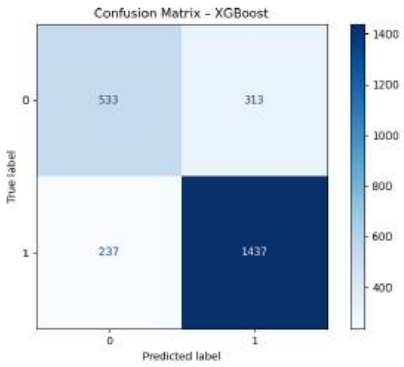



***Figure 8 : Confusion Matrix XGBoost Model***

***Comparative ROC Curve across Three Classification Models.***

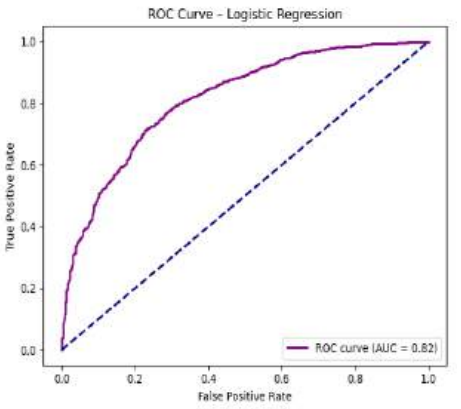



***Figure 9: ROC Curve for Social Assistance Eligibility Prediction – Logistic Regression (AUC = 0.82)***

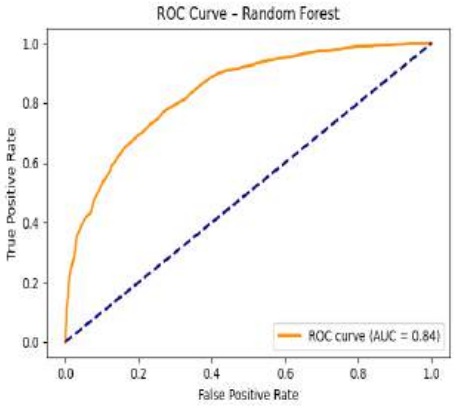



***Figure 10: ROC Curve for Social Assistance Eligibility Prediction – Random Forest (AUC = 0.84)***

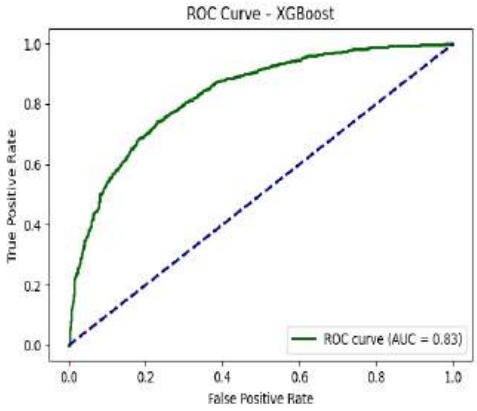



***Figure 11: ROC Curve for Social Assistance Eligibility Prediction – XGBoost (AUC = 0.83)***

Across the three confusion matrices, ensemble methods demonstrate improved classification accuracy and better error management. Both Random Forest and XGBoost show fewer false negatives, which indicates a higher capacity to correctly identify eligible individuals (Grover et al., 2020). Specifically, Random Forest achieves the highest true positive rate (1,470), followed by XGBoost (1,437), while Logistic Regression trails behind (1,429). Similarly, the lower false

ISSN: 2658-8455
Volume 6, Issue 9 (2025), pp. 303-318.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

positive rates in the ensemble models reflect better discrimination capacity and reduced misclassification.

Regarding the ROC curves, all models demonstrate strong discriminatory performance. Area Under the Curve (AUC) values range from 0.82 for Logistic Regression, to 0.83 for XGBoost, and 0.84 for Random Forest. An AUC above 0.80 is generally considered indicative of high-quality binary classification (Fawcett, 2006). These results confirm that tree-based ensemble models offer slightly superior ranking ability compared to the linear baseline.

Therefore, while Logistic Regression remains a valid and interpretable baseline, the ensemble models—Random Forest and XGBoost—demonstrate greater robustness, scalability, and predictive precision in handling socio-economic heterogeneity.

### 4.4. Hypotheses and Interpretation of Results

The empirical analysis also makes it possible to confront the findings with the hypotheses formulated in the literature review.

First, H1, which posited a positive relationship between educational attainment and the probability of being classified as eligible for social assistance, is clearly supported. Tree-based models consistently identify education_level as one of the most influential predictors. This result echoes previous studies emphasizing the decisive role of education in fostering socio-economic mobility and reducing poverty.

Regarding H2, which assumed that rural residence significantly increases eligibility, the results also confirm this hypothesis. The variable is_urban appears among the key determinants across the three models. This trend reflects the structural disadvantages faced by rural households, who are often less integrated into formal labor markets and have more limited access to basic services.

Finally, H3, which suggested that access to financial and digital services provides stronger predictive power than traditional demographic variables, is likewise validated. Predictors such as financial_inclusion, active_bank_user, and can_use_internet are highly ranked in both Random Forest and XGBoost models. Their relative weight sometimes surpasses that of conventional indicators such as age or marital status, underscoring the growing importance of financial and digital inclusion in shaping social vulnerability.

Taken together, these findings confirm the relevance of the three hypotheses. They highlight that education, geographic location, and access to financial and digital services are critical dimensions for understanding and predicting social assistance eligibility. These factors therefore need to be incorporated into any effort to renew targeting criteria, complementing traditional demographic and monetary indicators.

## 5. Ethical and Policy Considerations

To evaluate classification performance, we also conducted confusion matrices and ROC analysis of all three models (Logistic Regression, Random Forest, and XGBoost). They are complementary tools to assess how well the models discriminate between class 1 (eligible recipients) and class 0 (non-eligible recipients) over mere global accuracy measures (Hand & Till, 2001).

The use of machine learning for social assistance targeting offers many benefits, but also involves important ethical and governance considerations. They extend beyond algorithmic efficiency to include the values of transparency, accountability, and fairness in public decision-making (Jobin, Ienca & Vayena, 2019; OECD, 2022). A policy challenge in this era of predictive technology, the duty is clear: We cannot have that technology focus on outcomes that benefit some while sacrificing others.

One key ethical risk is bias in the algorithm. Training data that mirrors historical inequality or reflects structural disparities can cause machine learning models to encode discrimination. This is especially problematic in the domain of social protection where misclassification can result in unfair denial of life-saving benefits (Barocas, Hardt & Narayanan, 2019; Eubanks, 2018). For instance, marginalized communities are often under-counted or misrepresented because of biased measures, poor sampling, and legacy data systems.

Another challenge is the lack of interpretability in complex ML models—such as ensemble methods like Random Forest and XGBoost—which typically act as "black boxes." This lack of interpretability makes it difficult for the public to understand the basis for decision-making and can undermine institutional responsibility (Wachter, Mittelstadt, & Floridi, 2017). The "right to explanation" has been invoked as an aspirational quality of the governance of AI, emphasizing the importance of transparency in decision-making around social benefits eligibility. Explainable AI (XAI) methods, and tools like SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016) pertain to this idea and provide potential yet unexplored resources to improve the transparency of models to a great degree without much reduction in performance.

To address some of the abovementioned concerns, recent policy frameworks have provided ethical recommendations for the application of AI in social systems. For instance, the OECD (2022) suggests the use of interpretable models where possible, conducting regular algorithmic audits, and including end-users in the process of designing models. Such participatory methods aid in ensuring that technology-based solutions reflect the lived realities and indigenous rights of its benefactors (Veale & Binns, 2017).

Operationally, ethical use also concerns optimizing performance with social equity. Changing classification cutoffs in this way to minimize false negatives (at some cost in terms of false positives) is a normative decision based upon a principle of distributive justice (Hanna & Olken, 2018). These balancing acts must be made open, democratically discussed, and complemented with sound communication efforts to earn the public's trust.

In addition, alternative data sources (e.g., mobile phone metadata, browsing history, or geolocation) can enhance the predictive power, but they also raise new privacy and surveillance threats. This risk is particularly acute in low digital literacy areas or regions with poor data protection laws. Taking consent and restricting live data capture, only collecting data necessary for the purpose, and enforcing rigorous anonymization protocols are key in preserving integrity with regards to ethics (Athey, 2017; Chen et al., 2021).

In conclusion, ethical and responsible use of machine learning in social protection systems is not just about technical diligence, but also about institutional checks and balances, ethical foresight, and a social inclusion agenda. Predictive models need to be situated within a governance framework that respects human dignity, regulatory accountability, and deliberative policy-making. That is the only way, in turn, that algorithmic innovation can help in building fairer and more robust welfare systems.

Implementation, beyond these general principles, needs a protocol of treatment and a monitoring scheme. Its implementation can be staged initially, through pilot projects in certain areas to evaluate the integration of machine learning algorithms into current targeting systems. These pilots need to explore the tuning of eligibility cutoffs contextually to national priorities (especially the trade-off between minimizing exclusion and inclusion errors).

At the institutional level, capacity development is also required. This includes the training of program officers in how to make data-driven decisions, the building of dedicated data analytics units within social protection institutions, and specifically, the creation of a robust infrastructure for the stewardship of socio-economic data.

For the sake of accountability and ongoing improvement, deployment should be accompanied by a monitoring and evaluation (M&E) plan. KPIs can comprise therein one or more of the

ISSN: 2658-8455
Volume 6, Issue 9 (2025), pp. 303-318.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

errors of exclusion and inclusion, administrative cost savings, and beneficiary satisfaction. Annual updating of the models with new data would help to keep the predictions up to date while independent periodic audits as well as transparent reporting could bolster public confidence in the model.

This thus translates policy implications into an operational roadmap that connects algorithmic innovation to quantifiable gains in program effectiveness, fairness, and institutional accountability.

## 6. Conclusion and Perspectives

This research explored whether the performance of supervised machine learning models had the capacity to enhance the precision of targeting social protection programs among poor families. Using a synthetic dataset simulating realistic socio-economic profiles, the models were evaluated based on standard classification metrics and feature importance analysis. The results demonstrate that ensemble methods, particularly Random Forest, offer superior performance compared to linear models, especially in reducing exclusion errors—an essential policy concern in welfare systems.

The analysis also highlights the importance of moving beyond simple demographic or income-based eligibility criteria. Features such as education level, urban residence, digital access, and geographic location emerged as key determinants of poverty-based eligibility across all models, underscoring the multidimensionality of vulnerability and the need for more nuanced data-driven targeting approaches.

However, while the technical results are encouraging, this study acknowledges several limitations. First, the use of a synthetic dataset, although realistic, may not fully capture the complexity, diversity, and noise found in actual administrative or household data. Consequently, the generalizability of the findings to real-world policy settings remains constrained without empirical validation.

Second, the models were trained using default hyperparameters and evaluated using a simple train/test split, without applying cross-validation or fine-tuning techniques. This choice was made to ensure a fair baseline comparison, but future research should explore optimization methods to improve robustness and reliability.

Third, although feature importance was explored, the study did not employ advanced explainability tools such as SHAP (SHapley Additive exPlanations) or partial dependence plots, which could further illuminate how variables interact to influence model predictions—especially critical in contexts where transparency and accountability are paramount.

Fourth, ethical considerations were addressed conceptually but not empirically assessed. No stakeholder consultation or institutional validation was conducted, which limits the operational readiness of the proposed models. Given the sensitivity of welfare targeting, future work should incorporate participatory approaches and ethical audits to ensure fairness, legitimacy, and social acceptability.

Looking ahead, several avenues for future research can be identified. These include:

- Testing the models on real administrative datasets and evaluating their performance across different socio-economic contexts.
- Integrating real-time behavioral or geospatial indicators to improve responsiveness.
- Applying explainable AI techniques to enhance transparency and user trust.
- Co-developing algorithmic tools with policymakers and social workers to ensure alignment with local policy goals and ethical standards.

In sum, machine learning offers promising capabilities to improve the efficiency and inclusiveness of social assistance systems. Yet, their deployment must be informed by robust empirical validation and continuous methodological improvements. Future work should

therefore focus on testing these models on real administrative datasets, integrating explainable AI techniques, and refining them through participatory approaches with policymakers and social workers. In this way, predictive modeling can evolve from a proof of concept into a reliable decision-support tool for strengthening social protection systems.

## References:


(1). Aarvik, P. (2021). *Automated and accountable? Algorithmic targeting in social protection*. Norwegian Institute of International Affairs.

(2). Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. (2020). Roles for computing in social change. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAccT)*, 252–260.

(3). Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. (2022). Targeting development aid with mobile phone data: Evidence from an experiment in Afghanistan. *American Economic Review*, *112*(11), 3836–3871. https://doi.org/10.1257/aer.20211679

(4). Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science, 355*(6324), 483–485. https://doi.org/10.1126/science.aal4321

(5). Athey, S., & Imbens, G. (2019). Machine learning methods that economists should know about. *Annual Review of Economics, 11*, 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

(6). Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*. http://fairmlbook.org/

(7). Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

(8). Beegle, K., Coudouel, A., & Monsalve, E. (2018). *Realizing the full potential of social safety nets in Africa*. World Bank. https://doi.org/10.1596/978-1-4648-1164-0

(9). Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*(2), 281–305.

(10). Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test, 25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

(11). Binns, R. (2020). On the apparent conflict between individual and group fairness. *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 514–524. https://doi.org/10.1145/3351095.3372864

(12). Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

(13). Brown, C., & Ravallion, M. (2020). In praise of price subsidies. *Economics & Politics, 32*(3), 351–369. https://doi.org/10.1111/ecpo.12154

(14). Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

(15). Chen, X., Lin, L., & Chen, C. (2021). Leveraging synthetic data for machine learning fairness: A case study in healthcare. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 115–121. https://doi.org/10.1145/3461702.3462624

(16). Chouldechova, A., & Roth, A. (2020). A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM, 63*(5), 82–89. https://doi.org/10.1145/3376898

(17). Duarte, J., & Azevedo, J. P. (2022). Using machine learning to support social assistance targeting: Evidence from Latin America. *World Bank Policy Research Working Paper*

ISSN: 2658-8455
Volume 6, Issue 9 (2025), pp. 303-318.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

*No. 10046*. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/

(18). El Emam, K., Mosquera, L., Bassil, S., & Buckeridge, D. (2020). Evaluating identity disclosure risk in fully synthetic health data: Model development and validation. *Journal of Medical Internet Research, 22*(11), e23139. https://doi.org/10.2196/23139

(19). Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

(20). Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

(21). Gentilini, U., Almenfi, M., Orton, I., & Dale, P. (2022). *Social protection and jobs responses to COVID-19: A real-time review of country measures*. World Bank. https://openknowledge.worldbank.org/handle/10986/33635

(22). Goncalves, A., Ray, P., Soper, B., Stevens, J., Coyle, L., & Sales, A. P. (2020). Generation and evaluation of synthetic patient data. *BMC Medical Research Methodology, 20*, 108. https://doi.org/10.1186/s12874-020-00977-1

(23). Grover, A., Leskovec, J., & Guestrin, C. (2020). Fairness in decision-making: From models to policy. *Advances in Neural Information Processing Systems, 33*. https://papers.nips.cc/paper/2020/hash/

(24). Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning, 77*(1), 103–123. https://doi.org/10.1007/s10994-009-5119-5

(25). Hand, D. J., & Till, R. J. (2001). A simple generalization of the area under the ROC curve for multiple class classification problems. *Machine Learning, 45*(2), 171–186. https://doi.org/10.1023/A:1010920819831

(26). Hanna, R., & Olken, B. A. (2018). Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries. *Journal of Economic Perspectives, 32*(4), 201–226. https://doi.org/10.1257/jep.32.4.201

(27). Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

(28). Kidd, S., & Athias, D. (2020). *Hit and miss: An assessment of targeting effectiveness in social protection*. Development Pathways.

(29). Kim, J. Y. (2021). Poverty prediction with machine learning: Evidence from Costa Rica. *arXiv preprint*. arXiv:2111.13319. https://arxiv.org/abs/2111.13319

(30). Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–705. https://scholarship.law.upenn.edu/penn_law_review/

(31). Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*. https://papers.nips.cc/paper/2017/hash/

(32). Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable*. https://christophm.github.io/interpretable-ml-book/

(33). Narayanan, A., & Vallor, S. (2021). Why software engineering courses should include ethics coverage. *Communications of the ACM, 64*(8), 30–32. https://doi.org/10.1145/3464905

(34). OECD. (2022). *Recommendation on the ethical use of artificial intelligence in the public sector*. Organisation for Economic Co-operation and Development. https://www.oecd.org/

(35). Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International*

*Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

(36). Salvador, E. L. (2024). Comparative analysis of machine learning algorithms for poverty classification: Evidence from the Philippines. *arXiv preprint*. arXiv:2407.13061. https://arxiv.org/abs/2407.13061

(37). Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society, 4*(2), 1–17. https://doi.org/10.1177/2053951717743530

(38). Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, *7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005

(39). World Bank. (2022). *Social protection and labor systems: Resilience and recovery from COVID-19*. https://www.worldbank.org/

(40). Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. O'Reilly Media.