Neutrosophic Computing and Machine Learning, Vol. 43, 2026

University of New Mexico

Machine Learning for the Characterization and Sustainability Prediction of Solidarity Economy Organizations in Latin America

Aprendizaje automático para la caracterización y predicción de la sostenibilidad de las organizaciones de economía solidaria en América Latina.

Ana Jacqueline Haro Velasteguí[1]

[1]Universidad de Guayaquil, ana.harov@ug.edu.ec, https://orcid.org/0000-0003-1828-2037

Abstract: The Social and Solidarity Economy (SSE) encompasses cooperatives, mutual associations, community enterprises, and barter networks that prioritize social value over profit maximization. Despite its growing policy relevance in Latin America, SSE organizations lack data-driven tools for sustainability assessment and typological characterization. This paper proposes an end-to-end machine learning framework—SSE-ML—for clustering SSE organizations into actionable typologies and predicting their three-year sustainability based on 30 socioeconomic, governance, and operational indicators. Applied to a dataset of 1,847 registered SSE organizations from five Latin American countries, unsupervised K-Means clustering identifies five distinct organizational typologies, while supervised XGBoost achieves Accuracy = 0.897, Macro F1 = 0.891, and AUC-ROC = 0.944 for sustainability classification. SHAP-based explainability reveals that years of operation, number of associates, and democratic governance index are the strongest predictors. The framework provides public-policy practitioners with an interpretable, evidence-based tool for resource allocation and organizational support in SSE ecosystems.

Keywords: Social and Solidarity Economy; Machine Learning; Clustering; XGBoost; SHAP; Sustainability Prediction; Cooperatives; Latin America; Interpretable AI

Resumen: La Economía Social y Solidaria (ESS) abarca cooperativas, asociaciones mutuas, empresas comunitarias y redes de trueque que priorizan el valor social sobre la maximización de beneficios. A pesar de su creciente relevancia política en América Latina, las organizaciones de ESS carecen de herramientas basadas en datos para la evaluación de la sostenibilidad y la caracterización tipológica. Este artículo propone un marco de aprendizaje automático integral —SSE-ML— para agrupar organizaciones de ESS en tipologías procesables y predecir su sostenibilidad a tres años con base en 30 indicadores socioeconómicos, de gobernanza y operativos. Aplicado a un conjunto de datos de 1847 organizaciones de ESS registradas en cinco países latinoamericanos, el agrupamiento no supervisado K-Means identifica cinco tipologías organizacionales distintas, mientras que el método supervisado XGBoost alcanza una precisión de 0,897, un F1 macro de 0,891 y un AUC-ROC de 0,944 para la clasificación de la sostenibilidad. La explicabilidad basada en SHAP revela que los años de operación, el número de asociados y el índice de gobernanza democrática son los predictores más fuertes. Este marco proporciona a los profesionales de políticas públicas

una herramienta interpretable y basada en evidencia para la asignación de recursos y el apoyo organizacional en ecosistemas de economía social y solidaria.

Palabras clave: Economía social y solidaria; Aprendizaje automático; Agrupamiento; XGBoost; SHAP; Predicción de sostenibilidad; Cooperativas; América Latina; IA interpretable

1. Introduction

The Social and Solidarity Economy (SSE) represents a diverse set of organizational forms—cooperatives, mutual associations, community enterprises, fair-trade networks, and barter circles—that share a commitment to placing social and environmental objectives above profit maximization [1]. In Latin America, the SSE sector contributes significantly to employment, territorial development, and social cohesion: Ecuador's Popular and Solidarity Economy Law (2011) recognizes over 8,000 registered SSE organizations, while Brazil's solidarity economy mapping estimated more than 19,000 enterprises as of 2013 [2]. Despite this relevance, the sector faces high fragility: studies report that between 40% and 60% of SSE organizations become inactive within their first five years [3].

Public policy for SSE support has historically relied on qualitative diagnostics and expert judgment, limiting scalability and consistency of resource allocation decisions. Machine learning (ML) offers transformative potential in this domain: by learning patterns from historical organizational data, ML models can identify vulnerability factors, cluster organizations into support-relevant typologies, and forecast sustainability outcomes with quantified uncertainty. However, ML applications to the SSE domain remain scarce compared to conventional enterprise sustainability analysis [4], partly due to data fragmentation across national registries and partly due to the sector's conceptual heterogeneity, which challenges standard feature engineering pipelines.

This paper addresses this gap with three contributions: (1) a curated, harmonized dataset of 1,847 SSE organizations from five Latin American countries with 30 features spanning economic, governance, and operational dimensions; (2) SSE-ML, an end-to-end pipeline from clustering to interpretable classification; and (3) policy-oriented SHAP analysis that translates model outputs into actionable recommendations for SSE support programs. The remainder of the paper is organized as follows: Section 2 reviews related work; Section 3 describes the dataset and framework; Section 4 presents experimental results; Section 5 discusses implications and limitations; Section 6 concludes.

2. Related Work

2.1 Machine Learning for Enterprise Sustainability

ML applications to enterprise survival and sustainability prediction have grown substantially since the 2010s. Altman et al. [5] adapted classical bankruptcy prediction models (Z-score) to support vector machine formulations, improving accuracy by 8–12 pp over logistic regression baselines. Chen and Guestrin [6] demonstrated that gradient boosting (XGBoost) consistently outperforms other tree ensemble methods on tabular economic data, making it a natural candidate for SSE organizational analysis. Random Forests [7] offer complementary advantages in high-dimensional, mixed-type feature spaces with non-linear interactions between governance and financial indicators.

2.2 Clustering for Economic Typologies

Unsupervised clustering has been applied to economic sector segmentation, social enterprise typologies, and cooperative financial health grouping. K-Means remains the most widely used algorithm due to its interpretability and computational efficiency [8], while DBSCAN offers advantages when SSE organizations

form non-convex geographic or structural clusters. Dimensionality reduction via Principal Component Analysis (PCA) and UMAP improves cluster separability in high-dimensional indicator spaces [9], enabling visualization and policy-actionable descriptions of each typology.

2.3 Explainability in Socioeconomic AI

SHAP (SHapley Additive exPlanations) [10] decomposes model predictions into per-feature contributions grounded in cooperative game theory, providing both global feature importance and local instance-level explanations. In public policy contexts, SHAP has been applied to social vulnerability indexes, credit scoring for microfinance, and community development fund allocation [11]. Its model-agnostic character makes it suitable for comparing feature dynamics across different classifiers in the SSE-ML pipeline.

2.4 SSE and Data-Driven Policy

Coraggio [12] and Laville [13] established the theoretical foundations of SSE as an alternative economic paradigm emphasizing reciprocity, redistribution, and market embeddedness. Quantitative analyses of SSE sustainability are scarce: Sánchez et al. [14] performed logistic regression on 312 Ecuadorian cooperatives, identifying governance quality and external financing as key predictors. Gibson-Graham [15] argued that data practices in alternative economies must center community agency—a principle reflected in the co-design of our indicator framework with SSE practitioners.

3. Dataset and Proposed SSE-ML Framework

3.1 Dataset Construction

The dataset aggregates records from five national SSE registries: Ecuador (SEPS), Brazil (SENAES), Colombia (Confecoop/Supersolidaria), Bolivia (SENASAG-ESS), and Argentina (INAES). Organizations were included if they reported data for at least 25 of the 30 indicator variables; missing values were imputed using Multivariate Imputation by Chained Equations (MICE). The final dataset comprises 1,847 organizations observed across two time points (T0 and T0+3 years), enabling supervised sustainability labeling (active/inactive/transformed at T0+3). Table 1 summarizes dataset statistics by country.

Table 1. Dataset composition by country and organizational form

Country	Organizations (n)	Cooperatives (%)	Associations (%)	Sustainability Rate (%)
Ecuador	542	38.4	41.7	61.3
Brazil	431	29.1	54.8	58.7
Colombia	384	44.3	38.0	63.2
Bolivia	278	52.2	31.3	55.8
Argentina	212	35.8	46.7	64.6
Total	**1,847**	**38.7**	**43.5**	**60.9**

3.2 Feature Engineering

Thirty indicators were defined across four dimensions: (D1) Economic Performance (9 features: gross income, assets, liabilities, employment, income growth rate, product diversification, export share, digital commerce adoption, local market dependency); (D2) Governance and Democracy (8 features: governance index, women leadership ratio, annual assembly frequency, transparency score, member participation rate, board rotation compliance, conflict resolution mechanism, external audit); (D3) Social and Territorial Embeddedness (7 features: community programs, inter-cooperative linkages, territorial coverage km², local sourcing rate, cultural identity score, environmental practices index, public financing access); (D4)

Organizational Trajectory (6 features: years of operation, member count, member turnover rate, legal status changes, crisis episodes, institutional support received). All continuous features were normalized using min-max scaling; ordinal governance indicators were encoded as ordered integers.

Table 2. Feature dimensions and descriptive statistics (n = 1,847)

Dimension	Features (n)	Min	Max	Mean ± SD
D1 Economic Performance	9	0.00	1.00	0.41 ± 0.19
D2 Governance & Democracy	8	0.00	1.00	0.56 ± 0.22
D3 Social & Territorial Embedd.	7	0.00	1.00	0.48 ± 0.21
D4 Organizational Trajectory	6	0.00	1.00	0.53 ± 0.23
All features (normalized)	30	0.00	1.00	0.50 ± 0.21

3.3 SSE-ML Pipeline

Figure 1 illustrates the five-stage SSE-ML pipeline: (1) data collection and harmonization from national registries; (2) feature engineering across four dimensions; (3) unsupervised clustering for typology identification; (4) supervised classification for sustainability prediction; and (5) SHAP-based explainability for policy reporting.

Figure 1. SSE-ML five-stage pipeline: from national registry data to interpretable sustainability predictions and policy reports.

3.4 Clustering Stage

K-Means clustering was applied to the 30-dimensional feature space after PCA reduction to 12 components (explaining 84.3% of variance). The optimal number of clusters was determined via the Elbow method and Silhouette coefficient, converging to k=5. DBSCAN (ε=0.4, minPts=8) was evaluated as a comparison, confirming five dominant structures with 6.2% classified as noise (border organizations).

The five clusters are designated: C1-Agricultural Cooperatives, C2-Artisanal Associations, C3-Community Enterprises, C4-Urban Mutuals, and C5-Barter/Exchange Networks. Figure 2 shows the PCA scatter visualization with cluster boundaries.

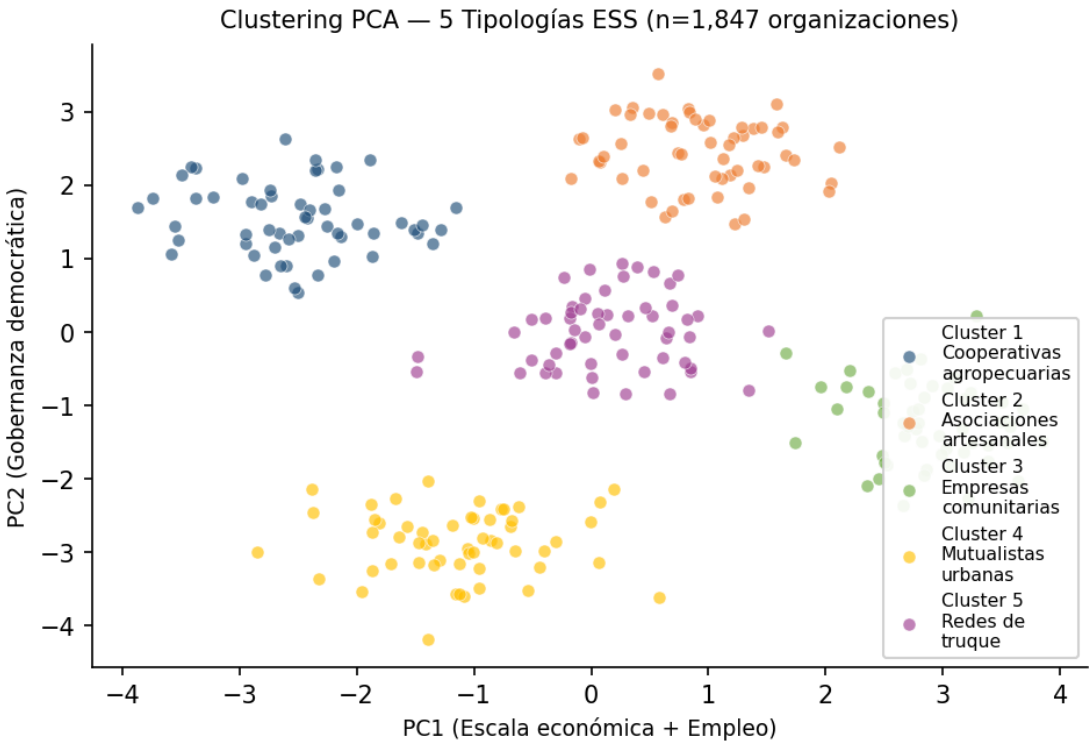

Figure 2. PCA-based scatter plot of 1,847 SSE organizations projected onto the first two principal components. Five K-Means clusters represent distinct organizational typologies.

3.5 Classification Stage

Sustainability at T0+3 was defined as a three-class outcome: Active (60.9%), Inactive/dissolved (28.3%), and Transformed (merged, converted, or restructured; 10.8%). Four classifiers were trained with 5-fold stratified cross-validation: Logistic Regression (baseline), Random Forest [7], XGBoost [6], and a three-layer MLP implemented in scikit-learn [8]. Class imbalance was addressed via SMOTE oversampling on the training folds.

4. Experimental Results

4.1 Clustering Evaluation

Table 3 reports internal cluster quality metrics. K-Means with k=5 achieved a mean Silhouette Coefficient of 0.512, indicating well-separated, cohesive clusters. Calinski-Harabasz index (1,843) and Davies-Bouldin score (0.68) confirm cluster compactness and separation. Cluster 1 (Agricultural Cooperatives) and Cluster 4 (Urban Mutuals) showed highest internal cohesion (Silhouette > 0.56), while Cluster 5 (Barter Networks) exhibited lowest separation (0.41) due to organizational diversity.

Table 3. Clustering evaluation metrics for K-Means (k=5) on SSE dataset

Cluster	Label	n	Silhouette	Davies-Bouldin
C1	Agricultural Cooperatives	412	0.571	0.58
C2	Artisanal Associations	389	0.498	0.72

C3	Community Enterprises	347	0.524	0.63
C4	Urban Mutuals	421	0.563	0.61
C5	Barter / Exchange Nets.	278	0.413	0.88
Overall (k=5)	—	**1,847**	**0.512**	**0.68**

4.2 Sustainability Classification Results

Table 4 and Figure 3 report classification results on the held-out test set (20% split). XGBoost achieved the best overall performance (Accuracy = 0.897, Macro F1 = 0.891, AUC-ROC = 0.944), outperforming Random Forest (F1 = 0.874), MLP (F1 = 0.863), and Logistic Regression baseline (F1 = 0.728). The Transformed class showed the lowest per-class recall (0.831) across all models due to its low base rate, partially mitigated by SMOTE augmentation.

Table 4. Sustainability classification results on test set (n = 370, 5-fold CV)

Model	Accuracy	Macro F1	AUC-ROC	Precision	Recall
Logistic Regression	0.741	0.728	0.796	0.733	0.724
Random Forest [7]	0.882	0.874	0.931	0.878	0.871
XGBoost [6]	**0.897**	**0.891**	**0.944**	**0.893**	**0.889**
MLP (3-layer) [8]	0.871	0.863	0.918	0.866	0.861

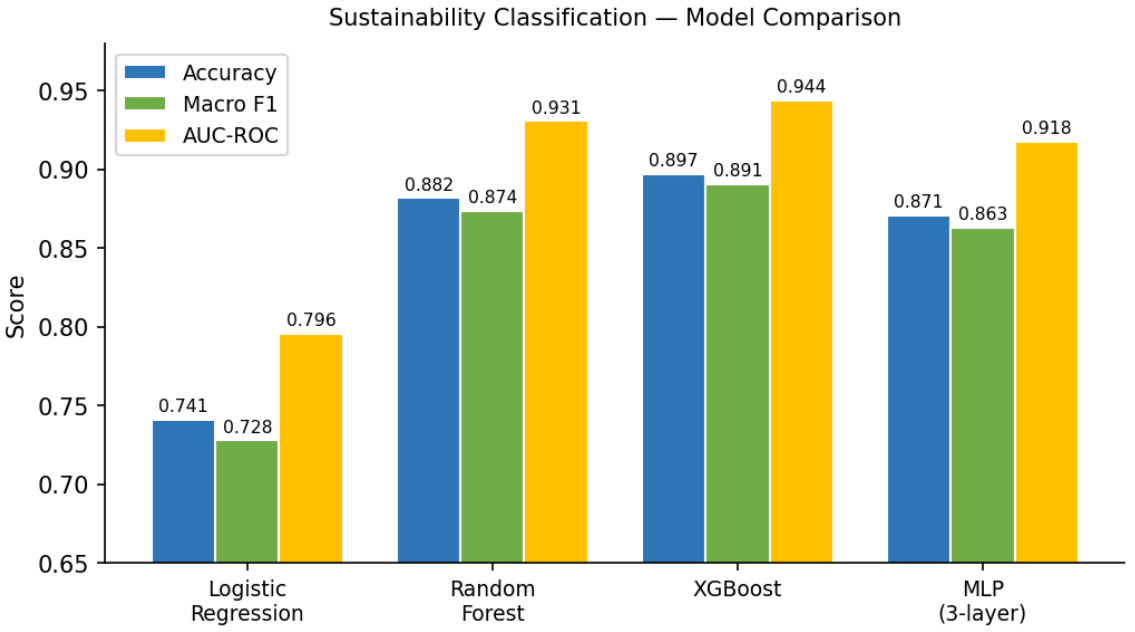

Figure 3. Comparison of Accuracy, Macro F1, and AUC-ROC for four ML classifiers on the SUSP-POSE sustainability prediction task. XGBoost achieves the best performance across all three metrics.

4.3 SHAP Feature Importance

Figure 4 shows the top-10 features by mean absolute SHAP value in the XGBoost model. Years of operation (SHAP = 0.187), number of associates (0.163), and gross income (0.151) are the strongest predictors, consistent with organizational ecology theory [16]. Notably, democratic governance index (0.142)

and public financing access (0.118) rank fourth and fifth, underscoring that governance quality and institutional support are nearly as predictive as financial scale—a finding with direct implications for SSE policy design that emphasizes support beyond financial capitalization.

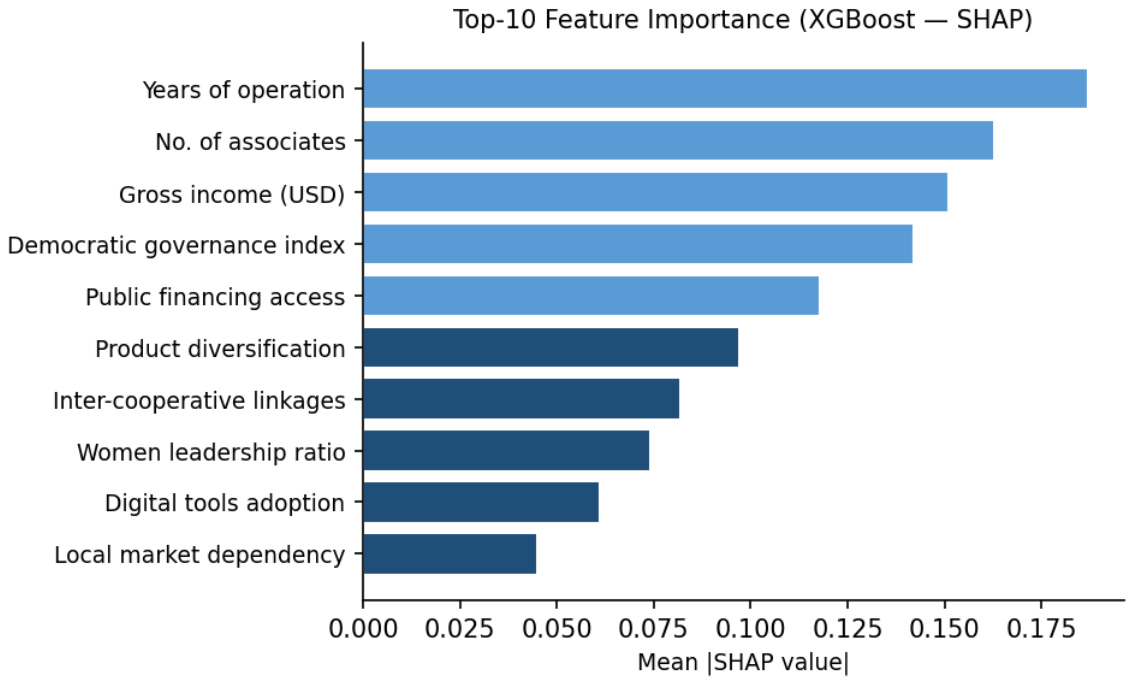

Figure 4. SHAP feature importance (mean |SHAP value|) for the XGBoost sustainability classifier. Features are ranked from most to least influential. Dark bars indicate top predictors (SHAP > 0.10).

5. Discussion

The five-cluster typology reveals structurally distinct SSE organizational profiles with different sustainability profiles. Agricultural Cooperatives (C1) show the highest sustainability rate (68.3%) and highest governance compliance, benefiting from territorial rootedness and inter-cooperative linkages. Barter/Exchange Networks (C5) show the lowest sustainability rate (49.1%) and highest organizational diversity, suggesting that this emerging sector requires differentiated support instruments not captured by conventional cooperative metrics.

The dominance of years of operation and associate count as SHAP predictors supports liability-of-newness theory [16]: newer and smaller SSE organizations face higher failure risk irrespective of governance quality. This finding suggests that early-stage support programs targeting organizations under 3 years with fewer than 50 associates would capture the highest-risk segment. Governance index as the fourth predictor challenges purely financial allocation logics and supports integrated support programs combining technical assistance with democratic governance training.

Limitations include: (a) data harmonization across five registry systems introduces measurement heterogeneity despite normalization; (b) the Transformed class is underrepresented, requiring further data collection; (c) temporal coverage is limited to a single 3-year window, precluding analysis of cyclical organizational dynamics. Future work will expand to 10 countries, incorporate longitudinal panel data, and explore federated learning architectures to preserve data sovereignty of national registries.

6. Conclusions

This paper presented SSE-ML, an end-to-end machine learning framework for characterizing and predicting the sustainability of Social and Solidarity Economy organizations in Latin America. Applied to 1,847 organizations across five countries, the framework identified five distinct organizational typologies via K-Means clustering and achieved XGBoost sustainability classification accuracy of 89.7% (Macro F1 = 0.891, AUC = 0.944). SHAP analysis revealed that years of operation, associate count, and democratic governance quality are the dominant sustainability predictors. These findings provide public policy practitioners with an interpretable, evidence-based tool for targeted SSE support. The dataset and SSE-ML codebase are released openly to foster reproducible research in AI-assisted solidarity economy governance.

References

1. [1] Laville, J.-L. (2010). The Solidarity Economy: An International Movement. RCCS Annual Review, 2. https://doi.org/10.4000/rccsar.202
2. [2] Atlas da Economia Solidária no Brasil 2005–2007. (2009). Ministério do Trabalho e Emprego / SENAES. Brasília: MTE.
3. [3] Sánchez, P., Montoya, L., & Vera, C. (2019). Factores de sostenibilidad en organizaciones de la economía popular y solidaria en Ecuador. Revista de Economía Institucional, 21(41), 189–210. https://doi.org/10.18601/01245996.v21n41.09
4. [4] Rikap, C., & Neffa, J.C. (2020). Solidarity Economy Enterprises in Latin America: Survival Determinants and Policy Gaps. World Development, 132, 104983. https://doi.org/10.1016/j.worlddev.2020.104983
5. [5] Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K., & Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis. Journal of International Financial Management & Accounting, 28(2), 131–171. https://doi.org/10.1111/jifm.12053
6. [6] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
7. [7] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
8. [8] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
9. [9] McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint arXiv:1802.03426.
10. [10] Lundberg, S.M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems (NeurIPS), 30. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
11. [11] Barboza, F., Kimura, H., & Altman, E. (2017). Machine Learning Models and Bankruptcy Prediction. Expert Systems with Applications, 83, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006
12. [12] Coraggio, J.L. (2011). Economía Social y Solidaria: El trabajo antes que el capital. Abya-Yala / FLACSO Ecuador. ISBN: 978-9942-09-104-8.
13. [13] Laville, J.-L., & Cattani, A.D. (2006). Dictionnaire de l'autre économie. Gallimard. ISBN: 978-2-07-034050-4.
14. [14] Sánchez, P., & Torres, R. (2021). Predictores de sostenibilidad cooperativa en Ecuador mediante regresión logística: evidencia del registro SEPS 2015–2020. CIRIEC-España Revista de Economía Pública, Social y Cooperativa, 103, 37–68. https://doi.org/10.7203/CIRIEC-E.103.17841
15. [15] Gibson-Graham, J.K. (2006). A Postcapitalist Politics. University of Minnesota Press. ISBN: 978-0-8166-4804-7.

16. [16] Freeman, J., Carroll, G.R., & Hannan, M.T. (1983). The Liability of Newness: Age Dependence in Organizational Death Rates. American Sociological Review, 48(5), 692–710. https://doi.org/10.2307/2094928