Definition of technological extension strategies based on exploratory analysis of the Sustainability Index using Artificial Intelligence: the case of oil palm producers in Colombia

Julián F. Becerra-Encinales[1,2]*, Brayan Rodríguez[2], Eloína Mesa-Fuquen[1]; Jorge A. Beltrán-Giraldo[1]; Alexandre P. Cooman[1], Paloma Bernal-Hernández[3], Luis H. Reyes[2], Juan C. Cruz[2]

[1]Colombian Oil Palm Research Centre Corporation, Cenipalma, Bogotá 111121, Colombia
Correspondence author *jbecerra@cenipalma.org and j.becerrae@uniandes.edu.co
[2]School of Engineering, Universidad de los Andes, Bogotá DC 111711, Colombia;
[3]University of Sussex, Science Policy Research Unit, SPRU, Brighton, BN1 9RH United Kingdom

INTRODUCTION

In the pursuit to foster technological adoption within the agricultural sector, rural extension services confront significant challenges (Anang et al., 2020), particularly in developing nations where technology adoption often lags (Ali Mengal et al., 2016). Within the realm of oil palm agro-industries, efficacious extension management frameworks are pivotal. Colombia's Oil Palm Research Center Corporation (Cenipalma) has gauged adoption rates through the Sustainability Index (SI), which considers economic, environmental, and social factors. This study undertakes an exploratory assessment of the SI in Colombian oil palm cultivation, aiming to identify technological adoption patterns and propose preliminary intervention strategies. Utilizing unsupervised machine learning, we discerned probable clusters (Morgenthaler, 2009). K-means and Ward clustering algorithms were employed to classify producers based on SI similarities (Garbely & Steiner, 2022). A nuanced characterization of each cluster was conducted, considering diverse variables such as land scale, palm oil zone, organizational model for fruit commercialization, cultivars, planting year, age group, and gender. Our findings provide valuable insights into the technology adoption patterns in Colombia's oil palm industry, offering a much more robust foundation for designing effective intervention strategies through data-driven approaches. As members of the agricultural community, it is crucial to recognize the significance of improving technology transfer and adoption to increase not only productivity but sustainability indexes.

METHODOLOGY

A conventional machine learning procedure was applied to the SI data, an endeavor akin to data mining, which laid the methodological groundwork for the initial project phase (Figure 1).

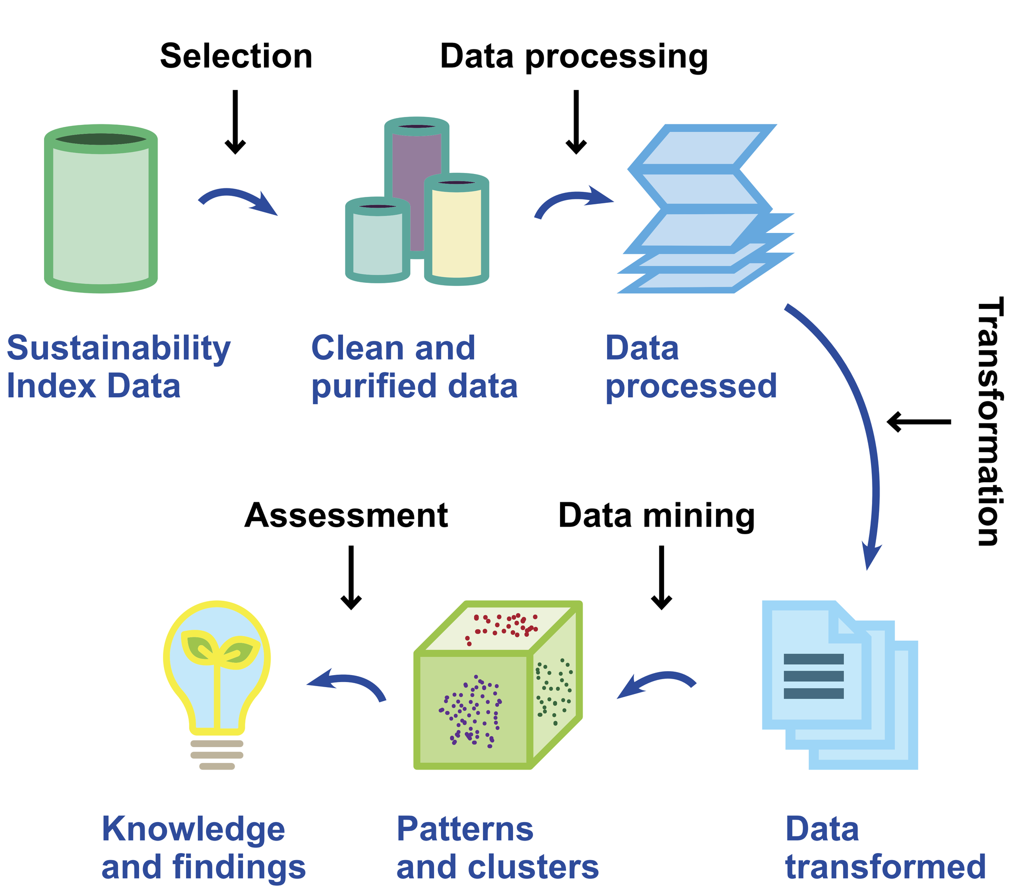

Figure 1. Schematic of the methodology for data mining. Source: Adapted from Yang et al., 2020

In the first stage, producer scoring data across established hierarchy levels (Table 1) was imported, followed by preprocessing tasks like data anonymization, duplicate record elimination, and producer classification based on age groups. Additional relevant data, including palm zone, producer scale, organizational model for fruit sale/purchase (palm nucleus), cultivars and age of the plantation were included.

Table 1. Sustainability Index database organized by hierarchy

Hierarchy	Amount of data
Producers with SI	3.808
Sustainability axes	3
Sustainability principles	10
Sustainability issues	29
Sustainability practices	79

Clustering algorithms, k-means and Ward's method, widely used for farmer group profile identification, were deployed on both the original scoring data across the 3 sustainability axes and the reduced-dimensionality data on 10 principles and 29 sustainability issues. Dimensionality levels elucidating roughly 70% of variance were chosen for the latter levels.

RESULTS AND CONCLUSIONS

The study examined the correlation between categorical variable correlations and their impact on producer classification. The findings revealed that sociodemographic variables emerged as significant cluster determinants, with Palm zone exerting the most substantial influence at the national level (Cramer's V=46.9%) (Figure 2), followed by Producer scale (Cramer's V=27.9%) and Age group (Cramer's V=17.5%). In contrast, Gender had a relatively lower impact (Cramer's V=10.0%). Producers were categorized into six clusters based on SI compliance and technology adoption levels (Figure 3).

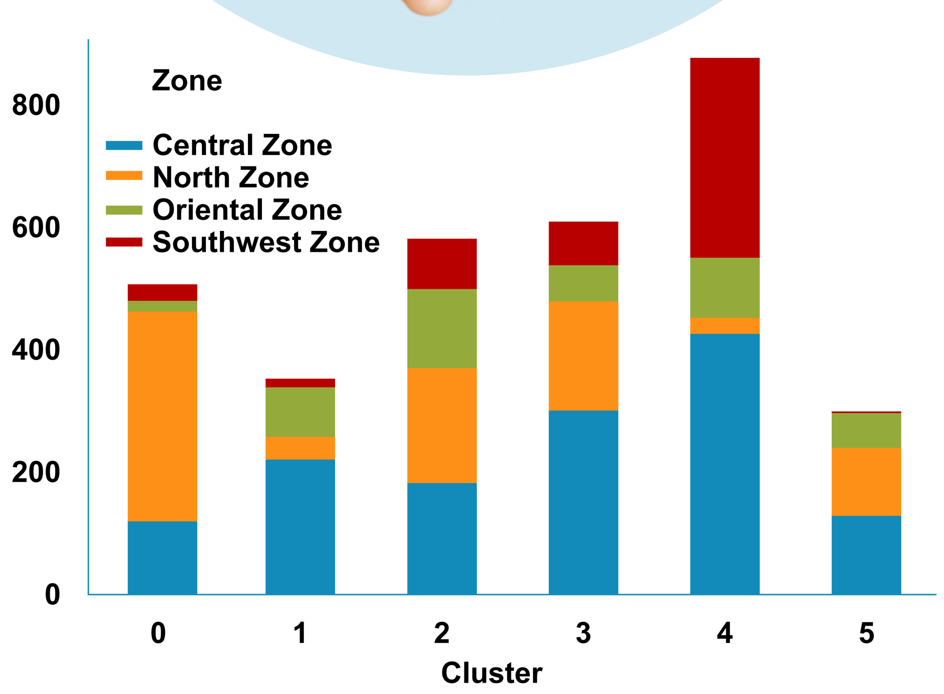

Figure 2. Distribution of clusters of producers classified according to levels of Sustainability Index (SI) compliance and technology adoption by palm zone. Analysis aided by Artificial Intelligence (AI)

The clustered outcomes were visualized in Cartesian planes, with each point representing a distinct producer. The environmental scores are shown along the X-axis, economic scores along the Y-axis, while the color gradient indicated the scores on the social axis. In front of each Cartesian plane, the distribution of producers in each cluster at the national level is shown on a heatmap. This visualization underscored a heightened efficacy in grouping at the sustainability axes level (Figure 3).

The crux of these findings pivots around an innovative classification paradigm for producers. Eschewing the traditional classification predicated on plantation size, this investigation introduces a nuanced approach considering producers' practices and technology adoption levels. Such an approach lays a robust foundation for subsequent research aimed at crafting producer archetypes and discerning the impact of various environments and contexts on technology adoption decision-making. To further refine insights, it is advocated to perform data mining on each Oil Palm zone independently, thereby unveiling more granulated patterns or regional groupings. This nuanced understanding, in turn, propels the formulation of precise and assertive technological extension strategies. In sum, this study accentuates the necessity of apprehending the sociodemographic variables that bear upon producer classification. By embracing a more holistic approach, agricultural cohorts are better positioned to devise efficacious strategies catering to the diverse needs across producer segments.

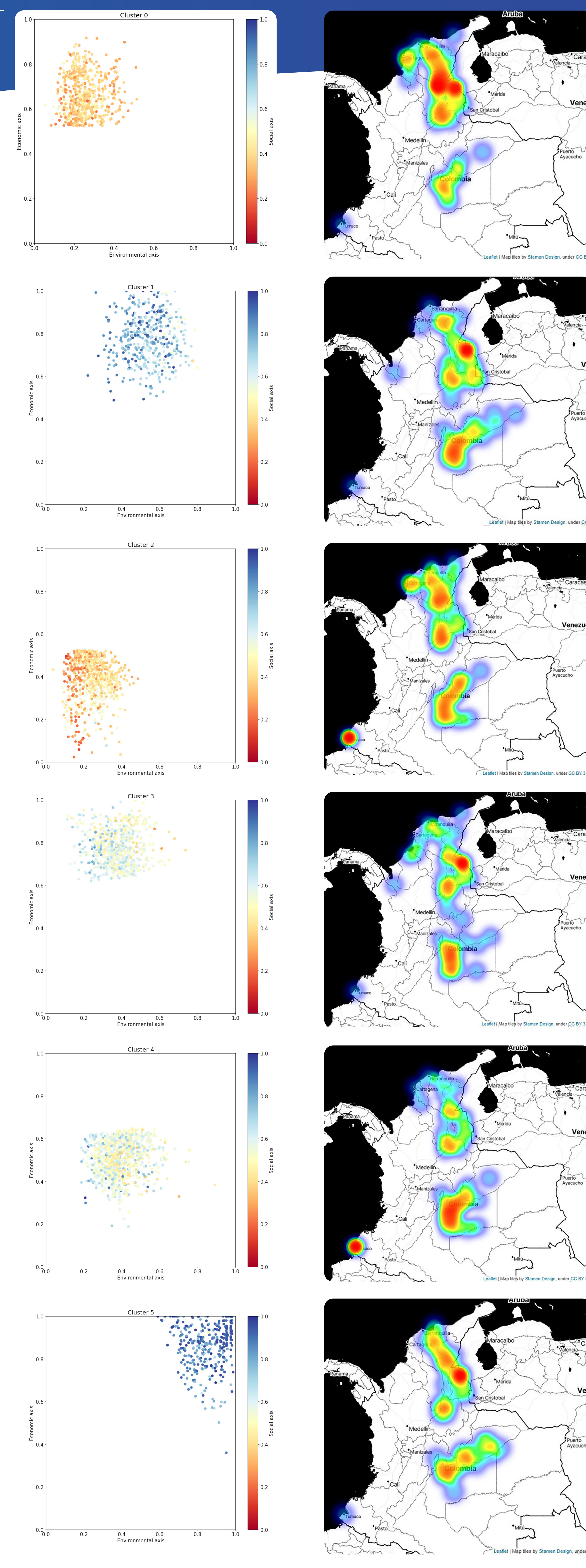

Figure 3. Geographical distribution, compliance and Technology Adoption Levels across Economic, Environmental, and Social Axes for Producers, clustered via Sustainability Index (SI) Analysis aided by Artificial Intelligence (AI)

REFERENCES

Ali Mengal, A., Habib, S., Baloch, F. M., Siddiqui, A. A., & Balochistan, Q. (2016). An innovative approach of public and private extension services regarding diffusion and adoption of agricultural technology in Balochistan, Pakistan: A conceptual framework. *JMAS The Journal of MacroTrends in Applied Science, 4*, 2016.

Anang, B. T., Bäckman, S., & Sipiläinen, T. (2020). Adoption and income effects of agricultural extension in northern Ghana. *Scientific African, 7*, 1–11. https://doi.org/10.1016/j.sciaf.2019.e00219

Garbely, A., & Steiner, E. (2022). Understanding compliance with voluntary sustainability standards: a machine learning approach. *Environment, Development and Sustainability.* https://doi.org/10.1007/s10668-022-02524-y

Morgenthaler, S. (2009). Exploratory data analysis. , 1(1), 33-44. *Wiley Interdisciplinary Reviews: Computational Statistics, 1*(1), 33–34.

Yang, A., Zhang, W., Wang, J., Yang, K., Han, Y., & Zhang, L. (2020). Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA. In *Frontiers in Bioengineering and Biotechnology* (Vol. 8). Frontiers Media S.A. https://doi.org/10.3389/fbioe.2020.01032