

Predicción de indicadores multidimensionales de la pobreza mediante aprendizaje automático: Un estudio con datos a nivel municipal en México

Análisis de Ciencia de Datos (TC2004B)

Mariel Álvarez Salas
*Escuela de Ingeniería y Ciencias
Tecnológico de Monterrey*
Monterrey, NL
A01198828@tec.mx

Maria Paula Recinos Rios
*Escuela de Ingeniería y Ciencias
Tecnológico de Monterrey*
Monterrey, NL
A00840131@tec.mx

Jorge Alfonso Andujo Valdez
*Escuela de Ingeniería y Ciencias
Tecnológico de Monterrey*
Monterrey, NL
A01563535@tec.mx

Valeria Arciga Valencia
*Escuela de Ingeniería y Ciencias
Tecnológico de Monterrey*
Monterrey, NL
A01737555@tec.mx

Ana Sofia Ibarra Ponce
*Escuela de Ingeniería y Ciencias
Tecnológico de Monterrey*
Monterrey, NL
A01285836@tec.mx

Resumen—El presente estudio evaluó la eficacia de 10 modelos de aprendizaje automático para predecir los niveles de pobreza en México, con el objetivo de identificar las principales áreas de oportunidad para invertir recursos y esfuerzos en la reducción del índice de pobreza.

El estudio analiza datos del Censo de Población y Vivienda 2020 del INEGI, y las variables objetivo pertenecen a la medición multidimensional de la pobreza 2020 del CONEVAL. Los resultados sugieren que los factores clave a considerar son la baja escolaridad, la desigualdad de género, el acceso a la tecnología y la falta de infraestructura en la vivienda.

Index Terms—Pobreza, aprendizaje automático, modelos predictivos, infraestructura, CONEVAL, INEGI, México

I. INTRODUCCIÓN

La pobreza en México es un fenómeno complejo que afecta a una proporción significativa de la población, impactando negativamente en su bienestar y limitando las oportunidades de desarrollo económico y social del país. Según el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) [13], en 2022, el 36.3 % de la población mexicana se encontraba en situación de pobreza, lo que equivale a aproximadamente 46.8 millones de personas. Aunque esta cifra representa una disminución en comparación con el 41.9 % registrado en 2018, la persistencia de altos niveles de pobreza sigue siendo un desafío crítico para el desarrollo nacional. La pobreza no solo implica la falta de ingresos suficientes, sino que también se manifiesta en múltiples carencias sociales, como el rezago educativo, el acceso limitado a servicios de salud, las deficiencias en servicios básicos de la vivienda, entre otras. Estas privaciones afectan la calidad de vida de millones de mexicanos, enfrentando así desigualdad y exclusión social.

En este contexto, es fundamental identificar y priorizar los sectores donde la inversión pública y privada puede tener un impacto más significativo en la reducción de la pobreza.

El *Instituto Nacional de Estadística y Geografía* (INEGI) es la institución encargada de recopilar y difundir información estadística y geográfica en México. Por su parte, el *Consejo Nacional de Evaluación de la Política de Desarrollo Social* (CONEVAL) es responsable de la medición del bienestar social y económico en distintos niveles de gobierno, a través de indicadores como el *Índice de Medición Multidimensional de la Pobreza*. Para este propósito, el CONEVAL utiliza información generada por el INEGI, principalmente en censos de población y vivienda.

El objetivo de esta investigación es desarrollar un modelo de regresión capaz de predecir los indicadores de Medición Multidimensional de la Pobreza generados por el CONEVAL, a partir de los datos del Censo de Población y Vivienda 2020 del INEGI a nivel municipal. Asimismo, se busca responder la siguiente pregunta de investigación: ¿En qué sectores debería priorizarse la inversión para reducir los indicadores de pobreza?

Los datos del INEGI contienen información sobre la cantidad de habitantes en el país, así como sus principales características demográficas, socioeconómicas y culturales. En el caso de los indicadores del CONEVAL, estos ofrecen una medición multidimensional de la pobreza, combinando información del INEGI con datos de la *Encuesta Nacional de Ingresos y Gastos de los Hogares* (ENIGH).

En esta investigación, se consideran dos indicadores clave como variables objetivo:

- Carencia por acceso a los servicios básicos en la vivien-

da: Representa la falta de infraestructura esencial.

- Población en situación de pobreza: Personas con al menos una carencia social y cuyos ingresos no les permiten adquirir bienes y servicios básicos.

II. REVISIÓN DE LITERATURA

La literatura actual del tema aborda diferentes enfoques metodológicos, conjuntos de datos y regiones geográficas. En cuanto al uso de algoritmos de boosting en la clasificación de la pobreza, un estudio en Filipinas [10] se identificó que CatBoost obtuvo el mejor rendimiento con un 91 % de precisión, mientras que AdaBoost presentó el menor desempeño. Este hallazgo subraya la importancia de seleccionar modelos que manejen eficientemente datos de alta dimensión y la necesidad de una adecuada limpieza y normalización de los datos para evitar sesgos. En el contexto de la pobreza relativa en áreas rurales, el uso de modelos como XGBoost, Random Forest y regresión logística permitió evaluar la influencia de distintas variables, donde XGBoost destacó con un 81 % de precisión, lo que sugiere que los métodos tradicionales aún pueden ofrecer aproximaciones sólidas en la clasificación de la pobreza.

La aplicación de modelos de clasificación en la predicción de la pobreza en Costa Rica [8] demostró que Gradient Boosted Trees y Random Forest fueron los más efectivos, alcanzando una precisión del 68.45 % y 64.9 %, respectivamente. Sin embargo, se señala que estos modelos podrían mejorar con un tratamiento más refinado de los datos faltantes. Un hallazgo clave es la confirmación de que la educación es un factor determinante en la predicción del nivel socioeconómico de los hogares, lo que refuerza la necesidad de integrar variables de dimensión educativa en los modelos predictivos.

Se notó que en la mayoría de investigaciones afirma al modelo 'Random Forest' como óptimo. La exploración de métodos de Random Forest multivariado [2] reveló que estos algoritmos son altamente efectivos para estimar la pobreza de distintos países en tiempo real, alcanzando coeficientes de determinación de hasta 0.7 en la estimación de la pobreza actual. Sin embargo, muestran limitaciones en la predicción de la pobreza futura, con coeficientes que oscilan entre 0.21 y 0.14, lo que indica que su aplicabilidad es más adecuada para el análisis de datos prescriptivos en lugar de pronósticos a largo plazo.

En una investigación más detallada a los distintos indicadores de pobreza elegidos, los artículos analizados abordan la predicción de *población en situación de pobreza* a través del uso de modelos de aprendizaje automático en distintos contextos y metodologías. En el caso de Costa Rica [12], el estudio utilizó datos de encuestas de hogares para predecir la pobreza mediante modelos de XGBoost, regresión múltiple y árboles de decisión, encontrando que los modelos basados en boosting ofrecieron los mejores resultados. Se implementaron diversas estrategias de selección de variables, destacando que eliminar demasiadas características afecta negativamente el rendimiento.

En el estudio de Kirguistán [9], se analizaron datos de la Demographic and Health Survey (DHS), donde se transformó el índice de riqueza en una variable binaria para clasificar los hogares en pobreza o no pobreza. Se probaron modelos como XGBoost y Modelos Lineales Generalizados (GLM), encontrando que XGBoost obtuvo la mayor precisión (AUC = 0.911). Un hallazgo clave es que, si bien agregar más información puede mejorar la predicción, una selección estratégica de variables puede reducir los costos computacionales. En conjunto, estos estudios resaltan que los modelos de boosting, en particular XGBoost, han demostrado ser altamente efectivos en la predicción de la pobreza al manejar datos de alta dimensionalidad. Un punto en común es que la selección adecuada de variables es clave para optimizar los modelos y que, aunque más información puede mejorar las predicciones, en algunos casos una reducción estratégica de variables mantiene el rendimiento.

Por otro lado, para la identificación de *carencias en el acceso a servicios básicos de en la vivienda*, en el caso de India [1], se utilizó la National Family Health Survey 2015–2016 para desarrollar un Índice de Necesidades Básicas Insatisfechas (UBN), el cual incluye dimensiones como hacinamiento, acceso a electricidad, agua potable, saneamiento, asistencia escolar y capacidad de subsistencia. Se evaluaron diversos algoritmos supervisados de Machine Learning para clasificar los hogares según este índice, destacando que Random Forest obtuvo la mejor precisión (81.01 %), mientras que Naïve Bayes tuvo el menor rendimiento (78.27 %). El estudio concluyó que el hacinamiento era la dimensión más destacada de la pobreza, es necesario dar prioridad a la inversión en viviendas adecuadas que incluyan aseos y otras necesidades básicas.

III. METODOLOGÍA

Para el desarrollo de este estudio, se empleó la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), ampliamente utilizada en proyectos de ciencia de datos debido a su enfoque estructurado en el análisis y modelado de datos. Este enfoque permitió una integración sistemática de los datos provenientes de fuentes gubernamentales abiertas, asegurando un tratamiento adecuado de los datos y una selección óptima de modelos de predicción.

El estudio se centró en la predicción de dos indicadores clave de pobreza en México, previamente mencionados, con el fin de desarrollar modelos de aprendizaje automático que permitan inferir estos indicadores a partir de datos censales.

III-A. Contexto y Fuente de datos

Los datos utilizados en este estudio provienen de dos fuentes oficiales de acceso público en México:

- Censo de Población y Vivienda 2020 del INEGI: Proporciona información detallada sobre la población, características demográficas, condiciones de vivienda y variables geográficas a nivel municipal.
- Medición Multidimensional de la Pobreza 2020 del CONEVAL: Contiene indicadores de pobreza desagregados por estado y municipio, permitiendo evaluar la carencia



Figura 1. Visualización de metodología CRISP.

de servicios básicos en la vivienda y el porcentaje de población en situación de pobreza.

Para asegurar la coherencia en la integración de ambas bases de datos, se realizó un cruce de información a nivel municipal, estableciendo una estructura de datos unificada en la que cada observación corresponde a un municipio y sus características socioeconómicas asociadas. El tamaño del dataset final consistió en 2469 registros y 282 columnas.

III-B. Procesamiento de Datos

El preprocesamiento de datos fue una etapa fundamental para garantizar la calidad y robustez de los modelos de predicción. Se realizó la imputación de valores faltantes utilizando la mediana, dado que las distribuciones de los indicadores presentaban distribuciones sesgadas y eran los únicos con valores faltantes.

Debido a que se tenían 278 atributos posibles (dejando a un lado los municipio, clave de entidad y las variables predictoras), se llevó a cabo una selección de variables mediante un modelo de *Random Forest*. De esta manera, se identificaron aquellas con mayor relevancia para la predicción de cada uno de los indicadores de pobreza.

- Para la predicción de la población en situación de pobreza, se seleccionaron diez variables clave agrupadas en diferentes áreas:
 - Salud: Población sin afiliación a servicios de salud.
 - Vivienda: Falta de acceso a computadora e internet, ausencia de vehículos motorizados, carencia de líneas telefónicas y número reducido de habitaciones en la vivienda.
 - Educación: Población femenina de 15 años y más con primaria incompleta o sin escolaridad.
 - Ocupación: Tasa de desempleo en la población general y masculina de 12 años y más.
 - Población total: Cantidad de niñas de 8 a 14 años.
- Para la predicción de la carencia por acceso a los servicios básicos en la vivienda, se identificaron las siguientes variables relevantes:

- Vivienda: Falta de línea telefónica o celular, ausencia de tecnologías de la información y comunicación (TIC), carencia de agua entubada, piso de tierra, falta de drenaje y ausencia total de bienes en la vivienda.
- Educación: Población de 15 años y más analfabeta, con énfasis en la población femenina.
- Lengua indígena: Población de 3 años y más que habla alguna lengua indígena y también español, incluyendo una variable específica para mujeres en esta condición.

Dado que la mayoría de las variables presentaban distribuciones sesgadas a la derecha y contenían valores atípicos, se aplicó una transformación logarítmica para mejorar la distribución de los datos. A continuación en las figuras 2 y 3 se muestra un ejemplo del efecto de la transformación en algunas de las variables seleccionadas.

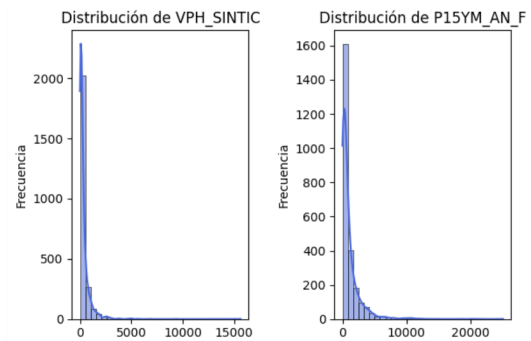


Figura 2. Distribución de variables antes de la transformación.

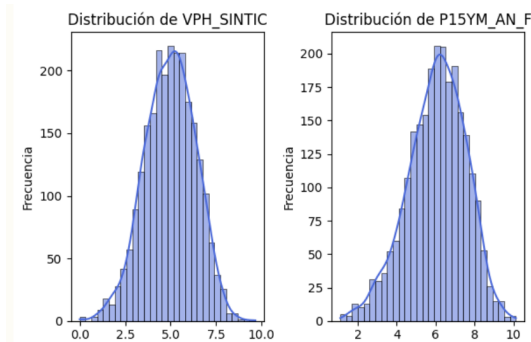


Figura 3. Distribución de variables después de la transformación.

En donde se observa que después de la transformación logarítmica [Figura 3] hay un cambio evidente en la distribución de los datos.

III-C. Modelos de Machine Learning

Para la predicción de los indicadores de pobreza, se evaluaron diversos modelos de aprendizaje automático, incluyendo métodos lineales, basados en árboles y enfoques de redes neuronales:

- Regresión Lineal (Linear Regression, Bayesian Ridge, Ridge)

- Modelos basados en árboles (Decision Tree, Random Forest, XGBoost, Gradient Boosting, AdaBoost)
- Modelos de Support Vector Regression (SVR)
- Redes Neuronales Artificiales (MLP Regressor)

Se seleccionaron 10 modelos para permitir una exploración amplia de los datos y seleccionar el modelo más adecuado para predecir cada indicador de pobreza, maximizando precisión y estabilidad sin comprometer interpretabilidad.

III-D. Indicadores de desempeño

Para evaluar el desempeño de los modelos de regresión, se utilizaron las métricas de error cuadrático medio (MSE) y coeficiente de determinación (R^2), la cuales permiten cuantificar tanto la precisión de las predicciones como la capacidad de los modelos para explicar la variabilidad en los datos.

La métrica de MSE cuantifica el error promedio al cuadrado entre las predicciones de un modelo y los valores reales. Esta es especialmente útil para detectar desviaciones significativas en las predicciones. Puede tomar valores desde 0 hasta el infinito positivo, siendo 0 una predicción perfecta por lo que se busca obtener valores de MSE pequeños. Por otro lado, el R^2 mide la proporción de la varianza total de la variable dependiente que es explicada por el modelo. Puede tomar valores desde 0 hasta 1, donde 1 significa que el modelo explica toda la variabilidad y, por tanto, es una predicción perfecta. Entonces, se busca obtener valores altos de R^2 .

III-E. Límites del estudio

Este estudio ha sido desarrollado a nivel municipal para todo el territorio nacional de México, por lo que existen limitaciones que es importante considerar al interpretar los resultados obtenidos.

En primer lugar, el análisis está basado en datos del INEGI que corresponden al año 2020, lo que implica un sesgo temporal significativo, considerando que los indicadores de pobreza y carencias sociales pueden cambiar en períodos de tiempo cortos, por lo que los modelos generados en el estudio pueden no reflejar las condiciones actuales. Esto resulta relevante considerando efectos de políticas públicas.

En cuanto a la selección de características, se empleó el algoritmo Random Forest para identificar las diez variables más relevantes para modelos predictivos. Este algoritmo puede verse influido por la multicolinealidad de variables. Además de que, al restringir el análisis a sólo diez variables, es posible una simplificación del fenómeno multidimensional de la pobreza o el acceso a servicios básicos dentro de la vivienda.

En cuanto a la modelación, si bien la variedad de algoritmos de regresión permitió capturar diversos patrones de relación entre variables y comparar su rendimiento; cada modelo posee ventajas y limitaciones propias que pueden influir en su desempeño. Se aplicaron técnicas para reducir riesgos de sobreajuste y evaluación en la división entre conjuntos de entrenamiento y prueba para entender el desempeño real de los modelos. Aun así, los resultados pueden depender de la calidad de datos y el ajuste, por lo que persiste cierta incertidumbre asociada al rendimiento general del modelo en otros contextos.

Finalmente, este estudio, al ser de enfoque cuantitativo, ya que no se han considerado factores territoriales que podrían enriquecer la comprensión del modelo, como pueden ser aspectos sociales, culturales o políticos que varían entre los municipios del país. Es relevante entender que esta limitación puede reducir la capacidad de los modelos para la captura de la complejidad estructural de la pobreza en cada municipio.

IV. RESULTADOS Y DISCUSIÓN

El procedimiento seguido para evaluar y encontrar el mejor modelo para cada indicador fue el siguiente: (1) evaluar cada modelo con los parámetros predeterminados; (2) evaluar cada modelo con parámetros predeterminados y validación cruzada con 10 folds; (3) evaluar cada modelo con los mejores hiperparámetros encontrados por una combinación de Randomized Search y Grid Search.

Para esto, y como se ha descrito anteriormente, en el presente estudio se evaluaron 10 modelos predictivos para cada uno de los indicadores. Sin embargo se realizó el ajuste de hiperparámetros y validación cruzada para dos de ellos, regresión lineal y ridge. En el cuadro 1, se presentan los resultados obtenidos para el indicador relacionado a la pobreza para dos de los modelos y siguiendo las fases del proceso de evaluación y en el cuadro 2 se presentan los resultados para el indicador relacionado a la carencia de acceso a servicios de la vivienda. Los otros modelos realizados fueron descartados con un análisis general que se incluye en los documentos dentro del repositorio en la sección *Disponibilidad de datos*.

Cuadro I
RESULTADOS DE MODELOS PARA INDICADOR POBREZA_POB

Métrica	Conjunto	Parámetros predeterminados	Validación Cruzada	GridSearch y Val. Cruzada
Regresión Lineal				
MSE	Train	0.005140	0.007388	0.007420
	Test	0.008496	0.003358	0.003385
R^2	Train	0.700803	0.628149	0.609923
	Test	0.564288	0.817282	0.777107
Ridge				
MSE	Train	0.005162	0.007415	0.007426
	Test	0.008558	0.003375	0.003378
R^2	Train	0.699484	0.626947	0.609639
	Test	0.561087	0.815468	0.777569

Cuadro II
RESULTADOS DE MODELOS PARA INDICADOR IC_SBV_POB

Métrica	Conjunto	Parámetros predeterminados	Validación Cruzada	GridSearch y Val. Cruzada
Regresión Lineal				
MSE	Train	0.006671	0.009028	0.008867
	Test	0.010356	0.005183	0.005196
R^2	Train	0.698705	0.623925	0.619690
	Test	0.565816	0.771779	0.753444
Ridge				
MSE	Train	0.006722	0.009077	0.008949
	Test	0.010373	0.005178	0.005160
R^2	Train	0.696392	0.621981	0.616178
	Test	0.565094	0.771529	0.755192

Considerando los resultados obtenidos, se realizó un análisis del desempeño con las métricas estadísticas de eficiencia y precisión descritas en la sección III-D, se ha seleccionado el modelo de regresión lineal múltiple como el que tiene un mejor desempeño y realiza mejores predicciones en las divisiones de entrenamiento y prueba sin realizar un sobreajuste, a diferencia del resto de modelos que fueron descartados desde un inicio debido a que tuvieron sobre ajuste aún con los mejores hiperparámetros seleccionados con gridsearch.

La simplicidad del modelo de regresión lineal resulta de gran importancia para su aplicación, ya que permite una interpretación directa de las variables significativas en los indicadores de medición multidimensional de la pobreza estudiados, lo que fomenta la transparencia y facilita el diseño de políticas con enfoque social.

Entre las variables predictivas que se seleccionaron para ambos indicadores, destacan las que están relacionadas con carencias educativas y tecnológicas, así como limitaciones materiales y habitacionales.

Estos resultados tienen relevancia al tener la capacidad de señalar áreas de prioridad en la intervención social, y enfatizan la necesidad de mejorar la infraestructura en viviendas.

Aunque existieron modelos con un desempeño similar al que se observa en la regresión lineal de los cuadros 1 y 2, su simplicidad interpretativa para su comunicación fueron clave y justifican la elección del modelo de regresión lineal por su desempeño consistente.

V. CONCLUSIONES

En este estudio, a través de la aplicación de técnicas de aprendizaje automático, se logró identificar las variables más relevantes en la estimación de dos indicadores clave: carencia por acceso a los servicios básicos en la vivienda y población en situación de pobreza.

Los resultados obtenidos permitieron analizar la relación entre factores socioeconómicos, educativos y de infraestructura con la incidencia de la pobreza, destacando la importancia de la vivienda, el acceso a servicios y las condiciones laborales como determinantes en los modelos predictivos. Asimismo, se identificaron desafíos metodológicos, como la presencia de multicolinealidad en ciertas variables, la necesidad de transformación de datos y la presencia de valores atípicos, los cuales fueron abordados mediante técnicas de preprocesamiento y selección de variables.

La aplicación de modelos de aprendizaje automático en el análisis de pobreza representa una herramienta poderosa para la toma de decisiones en políticas públicas, permitiendo dirigir esfuerzos y recursos hacia los sectores más vulnerables. Los hallazgos de esta investigación pueden contribuir al diseño de estrategias más efectivas para reducir las carencias sociales y mejorar la calidad de vida de la población en situación de pobreza.

Como trabajo futuro, se propone ampliar el análisis incorporando datos más recientes y explorando el impacto de factores económicos y políticos en los modelos predictivos. Además, el uso de enfoques más avanzados, como modelos de series

temporales o aprendizaje profundo, podría mejorar la precisión de las predicciones y proporcionar una visión más dinámica de la evolución de la pobreza en México.

En conclusión, este estudio demuestra el potencial del análisis de datos en la mitigación de la pobreza, proporcionando una base para futuras investigaciones y aplicaciones en el ámbito del desarrollo social.

VI. DISPONIBILIDAD DE DATOS

Con el objetivo de fomentar la transparencia y la mejora continua del presente estudio se pone a disposición del lector el repositorio con los datos procesados y *scripts* utilizados en el desarrollo del proyecto que se encuentra en el siguiente enlace.

<https://github.com/valeriaarciga/AnalysisDataScience>

El repositorio incluye:

- Scripts de preprocesamiento de datos
- Proceso de selección de variables
- Implementación de los modelos de regresión con procedimientos de validación cruzada y ajuste de hiperparámetros.

REFERENCIAS

- [1] Ansari, S., & Dhar, M. (2022). Poverty classification based on unsatisfied basic needs index: A comparison of supervised learning algorithms. *SN Social Sciences*, 2(5). Springer Nature. doi:10.1007/s43545-022-00375-y
- [2] Browne, C., Matteson, D. S., McBride, L., Hu, L., Liu, Y., Sun, Y., Wen, J., & Barrett, C. B. (2021). Multivariate random forest prediction of poverty and malnutrition prevalence. *PLoS ONE*, 16(9), e0255519. doi:10.1371/journal.pone.0255519
- [3] CONEVAL (s.f.). Medición de la Pobreza: Glosario. Recuperado de: <https://www.coneval.org.mx/Medicion/Paginas/Glosario.aspx#:~:text=Pobreza%3A...>
- [4] CONEVAL (s.f.). Medición de la Pobreza: Servicios básicos en la vivienda. Recuperado de: <https://www.coneval.org.mx/Medicion/Paginas/Medicion/Acceso-a-servicios-basicos-vivienda.aspx>
- [5] Gulecha, R. S., Reshmi, K. M., Rishitha, N., & Vani, K. (2024). Poverty Mapping in India using Machine Learning and Deep Learning Techniques. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4–2024, 319–326. doi:10.5194/isprs-annals-x-4-2024-319-2024
- [6] Hall, O., Dompae, F., Wahab, I., & Dzanku, F. M. (2023). A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications. *Journal of International Development*, 35(7), 1753–1768.
- [7] Huang, W., et al. (2023). What influences farmers' relative poverty in China: A global analysis based on statistical and interpretable machine learning methods. *Heliyon*, 9(9), e19525. doi:10.1016/j.heliyon.2023.e19525 [https://www.cell.com/heliyon/fulltext/S2405-8440\(23\)06733-6](https://www.cell.com/heliyon/fulltext/S2405-8440(23)06733-6)
- [8] Kim, J. Y. (2021). Using Machine Learning to Predict Poverty Status in Costa Rican Households. <https://research.ebsco.com/linkprocessor/plink?id=5a0ec49c-b7cc-36ca-8eb9-8101ace8c03b>
- [9] Li, Q., Yu, S., Échevin, D., & Fan, M. (2021). Is poverty predictable with machine learning? A study of DHS data from Kyrgyzstan. *Socio-Economic Planning Sciences*, 81, 101195. doi:10.1016/j.seps.2021.101195
- [10] Lynet, E. (2024). Use of Boosting Algorithms in Household-Level Poverty Measurement: A Machine Learning Approach to Predict and Classify Household Wealth Quintiles in the Philippines. <https://arxiv.org/abs/2407.13061>
- [11] Robb, K., Diaz Amigo, N., Marcoux, A., McAteer, M., & de Jong, J. (2022). Using Integrated City Data and Machine Learning to Identify and Intervene Early on Housing-Related Public Health Problems. *J Public Health Manag Pract*, 28(2), E497-E505. doi:10.1097/PHH.0000000000001343

- [12] Solís-Salazar, M., & Madrigal-Sanabria, J. (2022). Una propuesta de aprendizaje automático para predecir la pobreza. *Revista Tecnología En Marcha*, 35(4), 84–94. doi:10.18845/tm.v35i4.5766
- [13] Consejo Nacional de Evaluación de la Política de Desarrollo Social. (2023). Medición de la pobreza 2022: Comunicado de prensa. CONEVAL. Recuperado el 14 de marzo de 2025, de https://www.coneval.org.mx/SalaPrensa/Comunicadosprensa/Documents/2023/Comunicado_07_Medicion_Pobreza_2022.pdf

AGRADECIMIENTOS

Expresamos nuestro más sincero agradecimiento a la Dra. María de los Ángeles Constantino González, al Dr. Mauricio González Soto y al Dr. Rasikh Tariq por su invaluable orientación y apoyo durante el desarrollo de esta investigación. Sus comentarios, sugerencias y conocimientos fueron fundamentales para la mejora de este trabajo.

De igual forma, agradecemos a los líderes de parte del Socioformador INEGI, Elio Atenogenes Villaseñor García y a Ranyart Rodrigo Suarez Ponce de León, parte del laboratorio de Ciencia de Datos de la organización, por su colaboración y respaldo en distintas etapas del proyecto, lo que contribuyó al cumplimiento de los objetivos del trabajo.

Asimismo, agradecemos al Instituto Nacional de Estadística y Geografía (INEGI) y al Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) por la disponibilidad de los datos utilizados en este estudio.

Finalmente, extendemos nuestro reconocimiento a nuestros compañeros por su constante apoyo y motivación a lo largo de este proceso.