Artículo de Investigación

Aibi revista de investigación, administración e ingeniería. Volumen 14, Número 1, Pág 01-09, Enero - Abril de 2026, ISSN 2346-030X

Factores que inciden en la clasificación socioeconómica, usando redes neuronales para poblaciones rurales, Challhuahuacho, Cotabambas, Apurímac, Perú

Factors affecting socioeconomic classification using neural networks for rural populations: Challhuahuacho, Cotabambas, Apurímac, Peru

Roberto Eduardo Burga-Montañez[1]

[1]*Universidad Privada del Norte, Trujillo - Perú*

Recibido: 21 de agosto de 2025.
Aceptado: 11 de diciembre de 2025.
Publicado: 01 de enero de 2026.

Resumen- El APEIM, es una institución en el Perú, que ha logrado diseñar una propuesta metodológica de segmentación urbana en el Perú, sin embargo, las limitaciones respecto al estudio en las herramientas metodológicas, puede ayudar a crear una mejora significativa en las nuevas segmentaciones. Objetivo. Determinar que, factores influyen en la clasificación socioeconómica usando redes neuronales en poblaciones rurales, Challhuahuacho, Cotabambas, Apurímac, 2024. Método. La metodología fue cuantitativo, no experimental, transaccional, explicativa, usando información primaria a través de encuestas del INEI-ENAHO. La población fueron los jefes de hogar, específicamente del área rural del distrito de Challhuahuacho, provincia de Cotabambas, departamento de Apurímac. La técnica de procesamiento de datos fueron el uso de las Redes Neuronales Multicapa Perceptor, donde se particionaron tanto para el entrenamiento (80%), y otra para el modelamiento (20%), teniendo una arquitectura mínima de tres a 60 capas ocultas. Resultados. Entre los resultados se logró demostrar que existe diferencia significativa entre las poblaciones urbanas, aproximadamente el ochenta por ciento de la población del distrito de Challhuahuacho, la cual, posee nivel de estudios secundaria incompleta y menos de este grado educativo, por este motivo, utilizar la variable educación para la clasificación socioeconómica en ámbitos rurales sería poco relevante. Conclusión. Se concluye que, para segmentaciones rurales, el estudio logra identificar la necesidad de adaptar variables de clasificación a las características únicas de las áreas rurales.

Palabras clave: clasificación socioeconómica, redes neuronales, poblaciones rurales, JEL, Z13.

***Abstract*—** APEIM, an institution in Peru, has developed a methodological proposal for urban segmentation in Peru. However, limitations in the study of methodological tools could lead to significant improvements in new segmentations. Objective. To determine which factors influence socioeconomic classification using neural networks in rural populations in Challhuahuacho, Cotabambas, Apurímac, 2024. Method. The methodology was quantitative, non-experimental, cross-sectional, and explanatory, using primary data from INEI-ENAHO surveys. The population consisted of heads of household, specifically from the rural area of the district of Challhuahuacho, province of Cotabambas, department of Apurímac. The data processing technique used Perceptor Multilayer Neural Networks, partitioned for training (80%) and modeling (20%), with a minimum architecture of three to 60 hidden layers. Results. The results demonstrated a significant difference between urban and rural populations. Approximately eighty percent of the population of the Challhuahuacho district has incomplete secondary education or less. Therefore, using education as a socioeconomic variable in rural areas would be largely irrelevant. In conclusion, this study identifies the need to adapt classification variables to the unique characteristics of rural areas for rural segmentation.

Keywords: socioeconomic classification, neural networks, rural populations, JEL, Z13.

*Autor para correspondencia.
Correo electrónico: roberto.burga@upn.pe (Roberto Eduardo Burga Montañez).
La revisión por pares es responsabilidad de la Universidad de Santander.
Este es un artículo bajo la licencia CC BY (https://creativecommons.org/licenses/by/4.0/).
Como citar este artículo: R. E. Burga-Montañez, "Factores que inciden en la clasificación socioeconómica, usando redes neuronales para poblaciones rurales, Challhuahuacho, Cotabambas, Apurímac, Perú", Aibi revista de investigación, administración e ingeniería, vol. 14, no. 1, pp. 01-09 2026, doi: 10.15649/2346030X.5652

I. INTRODUCCIÓN

La clasificación socioeconómica es una herramienta que permite ordenar y priorizar estrategias de mercados, políticas económicas, como también aplicar una serie de eventos relacionados a la categorización de la sociedad en estratos, lo que resulta el análisis del acceso a la educación, salud, empleo y derechos entre los miembros, las cuales, esta desigualdad tiende a perdurar a lo largo del tiempo [1], tanto en generaciones, los cuales, se reduce el segmento, sea para la sociedad tenga mayor poder, así como en la influencia y acceso a bienes y servicios [1]. Sin embargo, el valor científico se encuentra en una metodología que permita segmentar de forma rural a las poblaciones, teniendo en cuenta los factores que se presenta en bases de datos conocidas. Si bien, la metodología utilizada en poblaciones con niveles de fragmentación es alta, existe una serie de alternativas que permitan incluir a estas poblaciones para obtener información primaria. Por ello, las metodologías en zonas rurales no han sido analizadas, debido a que, en la práctica, las empresas no les interesan, debido a la cantidad de fragmentación y por ser logísticamente difícil su aplicación.

Por este motivo, la atención en la investigación se enfoca tanto en la identificación en las variables socioeconómicas de mayor impacto, sea a través del proceso de estratificación, tanto en la población rural del distrito de Challhuahuacho, ubicado en la provincia de Cotabambas, departamento de Apurímac. La metodología de clasificación socioeconómica [1], fue establecida en el 2003 y ha sido aplicada consistentemente, planteada y reformulada por el APEIM [1], las cuales presenta limitaciones para el área rural, siendo un sesgo para objetivos orientados a estudios de mercado. Actualmente, la investigación, tiene sus limitaciones debido a ser un pueblo donde presentan condiciones propias, como es el caso de encontrarse cerca a un yacimiento petrolero, existen zonas rurales que no tengan las mismas condiciones, las cuales podrían sesgar los resultados, pero la metodología se mantendría igual [1].

La importancia de segmentar, permite a las empresas, activar estrategias de corto y largo plazo, para aplicar una estrategia de valor, aplicado a objetivos empresariales, sea a través de la comunicación como el marketing de marcas [1]. Esta aplicación, limita el accionar de las actuales empresas a reprocesar sus modelos tradicionales a nuevas metodologías de trabajo, utilizando herramientas más efectivas [1]. Por este motivo, la necesidad en el uso de modelos con IA, permitirían sistematizar sus procesos y adaptarlos en función a nuevos escenarios personalizados en función a los datos [2].

Debido a todo ello, el objetivo se basa en diseñar una propuesta metodológica de segmentación urbana en el Perú, sin embargo, las limitaciones respecto al estudio en las herramientas metodológicas, puede ayudar a crear una mejora significativa en las nuevas segmentaciones. Los resultados son de utilidad, tanto para los diferentes gobiernos central, regional y municipal, ya que permite ordenar una selección de posibles beneficiarios de programas sociales, abordar deficiencias en el proceso de estratificación social, particularmente en la filtración de usuarios que reciben programas sociales sin necesitarlos, lo cual, es una consecuencia de un proceso estratificación social deficiente [2].

II. METODOLOGÍA O PROCEDIMIENTOS

a. Antecedentes

Desde el 2001, el APEIM [2], ha logrado tener varias versiones para establecer mejoras con el cálculo de nivel socioeconómico, en base a la encuesta nacional de hogares, para la elaboración de pobreza, donde cada departamento está conformado por una distribución con la comparabilidad de datos, donde se encuentran registrados.

b. Justificación

Se justifica a nivel teórico, a que las empresas necesitan herramientas de trabajo para poder identificar a las personas, y lograr alguna evidencia empírica, respecto a la nueva segmentación social, basada en el desarrollo de una sociedad.

Si bien, las capacidades económicas (o monetaria) es una fuente de información veraz, de uso de herramientas tecnológicas con el fin de obtener información social, existen otras evidencias a las cuales no siempre se pueden tomar en cuenta, teniendo como precedente la aplicación de información como es el caso de la ENAHO.

c. Conceptos fundamentales

En el desarrollo del marco teórico, se ha vuelto una serie de conceptos relacionados a la aplicación, conceptos de niveles socioeconómicos, como teorías que respaldan dicha afirmación:

1. Teoría Socioeconómica

Se conceptualiza como el status socioeconómico que pueden ser clasificadas las personas, tal como la educación, nivel de ingreso (sueldo), ocupación, y el acceso a servicios básicos. [3], menciona que el status socioeconómico influye en el comportamiento individual y colectivo, determinando, tanto para el acceso a recursos materiales y sociales, por su parte, [4] argumentan la existencia de factores socioeconómicos que permite la integración en los individuos, donde se afecta directamente, tanto la calidad de vida como el acceso a servicios.

En un contexto rural, la existencia de la teoría socioeconómica, se aplica variables relacionadas al acceso a la educación, salud, servicios básicos y la vivienda, las cuales pueden influir en una clasificación socioeconómica, tanto para hogares rurales y urbanos [5].

2. *Teoría de Redes Neuronales*

Se consideran modelos computacionales inspirados en el funcionamiento del cerebro humano, donde se diseñan para identificar patrones y clasificar volúmenes de datos significativos. Por su parte, [6] describe cómo las redes neuronales pueden aproximarse a soluciones complejas

mediante el aprendizaje supervisado y no supervisado, lo que las convierte en herramientas efectivas para la clasificación y segmentación de datos. Estas redes se entrenan utilizando datos históricos para identificar patrones que luego se aplican a nuevos datos [7].

En este caso, las redes neuronales se emplean para clasificar a las poblaciones rurales en diferentes segmentos socioeconómicos utilizando diversas variables (como tamaño del hogar, acceso a servicios, ingresos, etc.) [8]. [9] el uso de redes neuronales, permite manejar información no lineal, para detectar patrones tanto en la clasificación de segmentos socioeconómicos que otros métodos tradicionales no logran identificar.

3. Teoría del Desarrollo Rural

Son características que se desarrollan en las áreas rurales, limitaciones estructurales y desigualdades en el acceso a recursos en comparación con las áreas urbanas. [10] enfatizaron el desarrollo rural, la cual, no solo implica la mejora de las condiciones económicas, sino también el acceso equitativo a servicios como la educación, salud y tecnología, elementos esenciales para mejorar la calidad de vida en estas áreas. [11] sostienen que la pobreza rural tiene un fuerte componente estructural relacionado por la falta de infraestructura y dependencia de recursos naturales.

Bajo este contexto, las poblaciones rurales y la teoría permiten comprender las desigualdades y limitaciones estructurales que enfrentan estas comunidades [12]. [13] en el análisis de políticas de desarrollo rural en América Latina, señala cómo la intervención pública en áreas rurales, a través de políticas de infraestructura y acceso a servicios, puede alterar las dinámicas socioeconómicas de estas comunidades.

4. Machine Learning aplicado a la clasificación socioeconómica

La técnica de Machine ayuda a la clasificación y predicción, en ese sentido, la identificación de patrones, permite crear modelos complejos sin depender de los tradicionales, en ese sentido, los métodos lineales no logran procesar grandes cantidades de información, como también su procedencia y de forma heterogénea, las cuales, el comportamiento en las poblaciones ha variado con el tiempo [12], en ese sentido, la comparación con métodos convencionales, han permitido acceso a variables, tales como vivienda, empleo, servicios, educación e infraestructura [15].

5. Deep Learning y redes neuronales profundas

La evolución de las redes neuronales tradicionales ha demostrado la composición de capas de aprendizaje, las cuales, se mantienen a través de arquitecturas complejas, las cuales se han caracterizado por las variables: sociales, económicas y demográficas, donde mejora el nivel predictivo, con indicadores de pobreza multi dimensional como son la detección de imágenes satelitales, censos y registros administrativos [15].

6. Desarrollo de la Hipótesis

Hipótesis 1 (H_1): No existen diferencias significativas en los factores y la clasificación socioeconómica en las poblaciones rurales de Challhuahuacho, Cotabambas, Apurímac.

Hipótesis Nula (H_0): Existen diferencias significativas en los factores y la clasificación socioeconómica en las poblaciones rurales de Challhuahuacho, Cotabambas, Apurímac.

III. METODOLOGÍA

La segmentación socioeconómica es considerada, como un proceso para entender las disparidades y dinámicas sociales dentro de una población, las investigaciones previas han identificado factores socioeconómicos que influyen en el comportamiento y las decisiones de las personas [14]. Por este motivo, la segmentación en poblaciones se encuentra en función de características socioeconómicas.

Bajo estas características, la metodología de investigación suele involucrar la identificación de variables clave que pueden utilizarse para distinguir diferentes grupos dentro de la población [15]. La adaptación respecto a los cuestionarios y la aplicación de técnicas estadísticas, permiten obtener conglomerados consistentes [16].

La población rural representa una parte significativa de la sociedad en muchos países y a menudo enfrenta desafíos socioeconómicos únicos, como es el caso de estudios previos, los cuales, han destacado las dinámicas que se tienen respecto a las comunidades rurales y a la necesidad de adaptar estrategias de investigación y políticas públicas a las necesidades particulares [17], además, la técnica de las redes neuronales han demostrado ser herramientas que se incluyen en la sociología y la economía. Estas técnicas de modelado computacional pueden capturar relaciones complejas entre variables y predecir comportamientos futuros con precisión [18].

Estos antecedentes proporcionan un contexto teórico y empírico para comprender la investigación sobre segmentación socioeconómica en poblaciones rurales [19]. Al contrastar los resultados obtenidos en el estudio con los hallazgos anteriores, se puede evaluar la consistencia y relevancia de los hallazgos actuales y generar nuevas líneas de investigación [20].

a. Diseño de investigación

[21] menciona que la estrategia de investigación fue utilizando información cuantitativa, debido a ello, y a la naturaleza de la investigación, la elección es de enfoque cuantitativo. En este contexto, el estudio se categoriza como no experimental, con el objetivo de evitar la manipulación deliberada de variables con el fin de obtener resultados específicos [22].

El diseño metodológico sigue la clasificación establecida por [23], que a través de un periodo de tiempo (transaccional), se encuentre la relación con el periodo de recolección de datos, donde esta se restringe a dos semanas. Utilizando el método de investigación explicativo [24],

detalla la evolución normal de eventos a investigar donde la relación causal entre la variable independiente y la dependiente (causa y efecto), se concluye que la investigación se clasifica como no experimental, ya que no implica la manipulación de hechos [24]. De acuerdo con [25], la información necesaria en la investigación demuestra, según propósitos, planteamiento del problema y objetivos, las características de una investigación con un enfoque cuantitativo.

A continuación, se muestra el modelo estadístico propuesto para la investigación:

$$Y_{(\text{cantidad de personas})} = \Omega X_1 + \beta X_2 + \lambda X_3 + \Psi X_4 + \varphi X_5 + \lambda X_6 + \theta X_7 + \mu X_8 + \alpha X_9$$

Donde:
Y= Cantidad de personas en el hogar
X_1 = Ocupación principal
X_2 =Tipo de vivienda
X_3 = Material de pisos
X_4 = Material de las paredes
X_5 = Dedicación al negocio
X_6 = Lugar de consulta para tratamiento de salud
X_7 = Dedicación del negocio
X_8 = Tenencia de terreno urbano
X_9 = Material de techos

Según la primera corrida, se consideró todas las variables de investigación, sin embargo, el programa seleccionó solo a nueve variables clave. El algoritmo tardó menos de dos segundos para el procesamiento de datos, así mismo, los errores calculados fueron el menor según el entrenamiento obtenido (Ver Tabla 1).

TABLA I
RESUMEN DEL MODELO

Entrenamiento	Error de suma de cuadrados	1,147
	Error relativo	,020
	Regla de parada utilizada	1 paso(s) consecutivo(s) sin disminución del error
	Tiempo de entrenamiento	0:00:00.03
Pruebas	Error de suma de cuadrados	,257
	Error relativo	,027

Con respecto a los resultados, se ha podido encontrar nueve variables (a diferencia de la clasificación urbana maneja cinco variables) que tendrían incidencia en el proceso de clasificación socioeconómica en la población rural del distrito de Challhuahuacho (Ver tabla 2).

TABLA II
INFORMACIÓN DE RED

Capa de entrada	Factores	1	Ud. se desempeñó en su ocupación principal o negocio, como:
		2	Tipo de vivienda (Por observación)
		3	El material predominante en los pisos es:
		4	El material predominante en las paredes exteriores es:
		5	¿A qué se dedica el negocio, organismo o empresa en la que trabajó?
		6	¿A dónde acudió para consultar recibir el tratamiento por este problema de salud?
		7	¿Cuál es la ocupación principal que desempeñó?
		8	En los últimos 12 meses, de……a…, ¿Usted o algún miembro de su hogar ha tenido o tiene algún terreno urbano?
		9	El material predominante en los techos es:
	Número de unidades [a]		50
Capas ocultas	Número de capas ocultas		1
	Número de unidades en la capa oculta		6
	Función de activación		Tangente hiperbólica
Capa de salida	Variables dependientes	1	¿Cuántas personas viven permanentemente en este hogar y están alojadas aquí?
	Número de unidades		1
	Método de cambio de escala para las dependientes de escala		Estandarizados
	Función de activación		Identidad
	Función de error		Suma de cuadrados

Según el diseño de la red neuronal, se ha visto que, según el grado de importancia de las variables, se establece la siguiente regla de clasificación (Ver tabla 3):

TABLA III
IMPORTANCIA DE LAS VARIABLES INDEPENDIENTES (PESOS PONDERADOS)

	Importancia	Importancia normalizada
Ud. se desempeñó en su ocupación principal o negocio, como:	,177	100,0%
Tipo de vivienda (Por observación)	,133	74,9%
El material predominante en los pisos es:	,096	54,2%
El material predominante en las paredes exteriores es:	,115	64,7%
¿A qué se dedica el negocio, organismo o empresa en la que trabajó?	,114	64,4%
¿A dónde acudió para consultar recibir el tratamiento por este problema de salud? (Acepte una o más alternativas)	,142	79,9%
¿Cuál es la ocupación principal que desempeñó? (Mostrar lista desplegable de la tabla de ocupaciones)	,080	45,0%
En los últimos 12 meses, de…….a…, Usted o algún miembro de su hogar ha tenido o tiene algún terreno urbano?	,057	32,0%
El material predominante en los techos es:	,086	48,5%

b. Población y muestra

La población rural se compone de 2,028 jefes de familia, según el censo de 2017, cuya información se obtiene a través del sitio web del INEI. Se aclara que la muestra adoptada sigue un enfoque probabilístico, conforme a la perspectiva de Hernández-Sampieri y Mendoza [26], basado en la selección de elementos dependiente de la probabilidad. Los instrumentos utilizados para la recolección de datos incluyen la adaptación del cuestionario de ENAHO. Además, el procesamiento de la información se llevó a cabo mediante el software estadístico SPSS versión 25.0.

Reemplazando: n= ((1.64)20.50.52028) / ((0.065)2(2028-1) +(1.64)20.50.5) =147.63≈148
Suponiendo que exista un alto nivel de incertidumbre (0,5=p)

La población de investigación fue los jefes de Hogar, del área rural del distrito de Challhuahuacho, provincia de Cotabambas, departamento de Apurímac.

El procedimiento de recolección de datos se llevó a cabo de la siguiente manera:

a. Población y Muestra: La población rural fue de 2,028 jefes de familia, donde según el censo del 2017 del INEI. Se utilizó un enfoque probabilístico para la selección de la muestra, conforme a la metodológica propuesta por [25]. La fórmula para calcular el tamaño de muestra se basó en un nivel de confianza del 95% y un margen de error del 5%. El tamaño de muestra calculado fue de aproximadamente 148 individuos.
b. Instrumentos y Técnicas de Recolección de Datos: Se empleó el cuestionario de la ENAHO para la recolección de datos. Este cuestionario fue adaptado para ajustarse a las características específicas de la población rural del distrito de Challhuahuacho, provincia de Cotabambas, departamento de Apurímac. La técnica utilizada para el procesamiento de datos fue la de redes neuronales multicapa perceptrón.
c. Software Utilizado: Una vez recopilados los datos, se procedió al procesamiento de la información mediante el software estadístico SPSS versión 25.0. Este software se utilizó para realizar análisis descriptivos, correlacionales y predictivos de los datos recopilados.

c. Procesamiento de datos

Para mejorar el análisis de procesamiento de datos, se realizó un análisis de limpieza en la base de datos, teniendo en cuenta, valores faltantes, críticos (no admitidos dentro de los patrones) y se considero solo aquellos, donde la base pueda ser analizada. Luego, se construyo un modelo estadístico, basado en redes neuronales, donde el 70% de la base de datos permite construir el modelo y el 30% fue para validarlo.

En la tabla 4, se observa la partición de datos que fue previamente declarada por el algoritmo para el entrenamiento y modelamiento:

TABLA IV
RESUMEN DE LA PARTICIÓN DEL ENTRENAMIENTO

Partición	Numero relativo	%
Entrenamiento	8	80
Prueba	2	20
Reserva	0	0
Total	10	100

Para poder construir el modelamiento, requiere el uso de capas ocultas, es decir, de pre procesamientos que permitan mejorar la predicción en el análisis de datos, teniendo un enfoque metodológico, como es el caso del uso máximo de capas ocultas (60). A través de la siguiente forma, se ha considerado la arquitectura del programa (Ver tabla 5):

TABLA V
RESUMEN DE LA PARTICIÓN DEL ENTRENAMIENTO

Número mínimo de unidades de la capa oculta	3
Número máximo de unidades de la capa oculta	60

De la misma forma, en la tabla 6, se observa el resumen del entrenamiento de la red neuronal, para establecer el nivel de aprendizaje:

TABLA VI
RESUMEN DEL ENTRENAMIENTO DE LA RED NEURONAL

Opción	Valor
Tasa de aprendizaje inicial	0,37
Límite inferior de la tasa aprendizaje	0.0001
Reducción de la tasa de aprendizaje	10
Impulso	0.9
Centro de intervalo	0
Desplazamiento del intervalo	±0,5

En la tabla 6, se observa que la tasa inicial de aprendizaje es de 0.37, es decir, que la rapidez con que se aplica en los procedimientos de la red neuronal, donde permite determinar el nivel de consistencia de los datos, debido a la limpieza anterior realizada. En el procesamiento de casos, se pudo observar, que el programa ha diseñado un 79,3% de entrenamiento y un 20.7% para las pruebas de diseños (Ver tabla 7).

TABLA VII
RESUMEN DE PROCESAMIENTO DE CASOS

		N	Porcentaje
Muestra	Entrenamiento	115	79,3%
	Pruebas	30	20,7%
Válido		145	100,0%
Excluido		3	
Total		148	

IV. RESULTADOS, ANÁLISIS E INTERPRETACIÓN

Las familias poseen mayor cantidad de miembros de familia tienen mayor capacidad de ingresos, se puede segmentar a un mayor nivel socioeconómico, teniendo en cuenta una condición para establecer niveles socioeconómicos (variable dependiente). Se trabaja con esta variable debido a que existe poca probabilidad en que las personas declaren sus ingresos con veracidad. Esta forma de desarrollar la clasificación, también lo realiza el APEIM para zonas urbanas.

Se utiliza la técnica de perceptrón multicapa de redes neuronales, para definir los factores que influyen en la cantidad de personas que integra una familia, esto nos permite clasificar el nivel socioeconómico de una familia. A efecto, que la técnica de clasificación, llamada clúster, utiliza algún tipo de segmentación en función a similitudes entre factores, con la finalidad de encontrar la dependencia entre las variables, argumentando la herramienta de redes neuronales en el diseño de los datos.

En la primera etapa de procesamiento de datos, se utilizó la técnica de redes neuronales para encontrar coeficientes (pesos) de los factores socioeconómicos, que tienen incidencia en la cantidad de personas que conforman el hogar. A continuación, se conceptualiza el modelo estadístico:

Y= Ocupación principal (0.177) +Tipo de vivienda (0,133) +material de pisos (0,096) +material de las paredes (0.115) +Dedicación al negocio (0.114)+ lugar de consulta para tratamiento de salud (0,142)+Dedicación del negocio (0.080)+tenencia de terreno urbano (0,057)+material de techos (0.086).

4.1 Efectividad del modelo empírico

La efectividad en la función, se realiza una simulación con la formula estadística (pronóstico de la variable) y la variable, con la finalidad de encontrar la cantidad de fallas. Usando la tabulación cruzada, se puede determinar que existe un 2,06% de error, es decir que existe un 97.93% de efectividad en la predicción de datos (Ver tabla 8).

TABLA VIII
TABLA CRUZADA ENTRE LA VARIABLE PRONOSTICADA Y LA REAL.

Recuento		Valor							Total
		1,00	2,00	3,00	4,00	5,00	6,00	7,00	
¿Cuántas personas Residen de manera continua en esta residencia y tienen su alojamiento en este lugar?	1	6	0	0	0	0	0	0	6
	2	0	24	0	0	0	0	0	24
	3	0	0	44	0	0	0	0	44
	4	0	0	0	18	3	0	0	21
	5	0	0	0	3	19	0	0	22
	6	0	0	0	0	3	16	0	19
	7	0	0	0	0	0	0	9	9
Total		6	24	44	21	25	16	9	145

TABLA IX
PERCENTILES DE LA VARIABLE DEPENDIENTE (NÚMERO DE INDIVIDUOS QUE RESIDEN EN LA RESIDENCIA).

¿Cuántas personas viven permanentemente en este hogar y están alojadas aquí?		
N	Válido	148
	Perdidos	0
Media		3,8649
Percentiles	25	3,0000
	50	3,5000
	75	5,0000

En base al criterio de percentiles, se propone tres niveles de clasificación (niveles socioeconómicos), según los resultados de los percentiles, se encontraría de la siguiente manera (Ver tabla 10):

TABLA X
PROPUESTA DE CLASIFICACIÓN DE CINCO NIVELES SOCIOECONÓMICOS SEGÚN LA CANTIDAD DE PERSONAS QUE VIVEN EN EL HOGAR

Nivel socioeconómico	Percentil	Conteo	Frecuencia	%
Nivel (C) Malo	25	1-2 personas	32	21.62%
Nivel (B) Regular	50	3-4 personas	65	43.92%
Nivel (A) Bueno	75	5-7 personas	51	34.46%

Discusión

Respecto a la discusión, los resultados ofrecen evidencias:

1. Los resultados revelan cuatro factores que influyen en la estratificación social: Ocupación principal, tipo de vivienda, nivel educativo y el género. Si bien, la estructura social y económica en las comunidades rurales, se basa en la interconexión social, [21] y [22] han demostrado que bienestar en las poblaciones tienen impactos positivos respecto a la capacidad de mantener algún sistema de información que permita ser interpretado como un desarrollo social.
2. Respecto a la metodología utilizada en la segmentación socioeconómica, [23] y [24] han utilizado enfoques de investigación, por conglomerados debido a la cantidad de variables que esta se encuentra al momento de clasificarla, sin embargo, en el caso de las redes neuronales, estas se pueden estratificar en función a contextos urbanos, como es el caso de requerir ajustes significativos en el entorno rural debido a la distribución de variables socioeconómicas y culturales [25]. Bajo Este punto resalta la necesidad de desarrollar y validar metodologías que sean sensibles a las particularidades de las poblaciones rurales [26].
3. Si bien, el objetivo se enfoca en determinar los factores, y durante todo el desarrollo se busca la clasificación socioeconómica, y al final se predice un número de personas por hogar, estas tres acciones, están dirigidas a inferir sobre la calidad de información que se obtiene al nivel socioeconómico, donde se justifica en mantener una serie de propuestas en el modelamiento estadístico para su aplicación técnica y teórica.
4. Los estudios evidencian que la clasificación socioeconómica en poblaciones rurales se tiene a través de variables que metodológicamente han sido diseñados en una ocupación, tipo de vivienda, materiales y acceso a servicios de salud, estas se encuentran a factores donde modifican el comportamiento de los indicadores que se encuentran asociados a la adaptación de los tradicionales.
5. Para el caso, Machine Learning, concide con [23], donde ha logrado integrar resultados basados en la clasificación y agrupamiento, donde se identifican niveles de pobreza en comunidades rurales. Esta formulación, en el uso de aprendizaje automático con técnicas estadísticas tradicionales, han permitido que el análisis no lineal entre los indicadores socioeconómicos, facilitan la focalización de Políticas Públicas.
6. La creciente evidencia respecto al reconocimiento de la inteligencia artificial, permiten inferir sobre condiciones socioeconómicas, siendo limitada a información oficial como el caso de patrones espaciales, donde la tendencia demuestra procesos de clasificación socioeconómica donde se analiza la predicción de pobreza rural [24], en el caso de investigaciones recientes tanto en la informalidad como en la sub declaración.
7. Para [1], la comparación entre las zonas urbanas y rurales (como es el caso de la propuesta), han permitido determinar acciones de contribución hacia la extrapolación de metodologías urbanas como territorios rurales, generando errores de clasificación con metodologías urbanas.

V. CONCLUSIONES

Se identificaron factores que influyen en la segmentación socioeconómica, como la ocupación principal, el tipo de vivienda, el material de pisos y paredes, la dedicación al negocio, el lugar de consulta para tratamiento de salud, la tenencia de terreno urbano y el material de techos. Estos hallazgos ofrecen una comprensión más profunda de cómo estos factores contribuyen a la estratificación socioeconómica de la población estudiada.

La existencia de una metodología de segmentación socioeconómica aplicada por el INEI y APEIM, que permite la determinación de una serie de atributos representativos. Este hallazgo sugiere que estas metodologías son fundamentales para comprender las características socioeconómicas de las poblaciones urbanas, lo que puede tener implicaciones importantes para la formulación de políticas y programas dirigidos a estas comunidades. Además, evidenciamos la importancia del bilingüismo en las entrevistas, resaltando la necesidad de considerar la diversidad lingüística en investigaciones socioeconómicas. Asimismo, proporcionamos información sobre el tamaño promedio de los hogares en zonas rurales, la participación diferencial de género en las entrevistas y la convivencia en hogares, lo cual puede informar políticas relacionadas con la vivienda, el bienestar familiar y la equidad de género.

VI. RECOMENDACIONES

Se identifica limitaciones tanto en la clasificación socioeconómica para zonas rurales, como al bajo nivel educativo, lo que destaca la necesidad de adaptar las metodologías de segmentación socioeconómica para tener en cuenta las características específicas de estas áreas. La investigación tiene una componente computacional, un modelo de redes neuronales, la capacidad de esta técnica estadística, es el auto aprendizaje, teniendo un entrenamiento de información y una selección para las pruebas de corrección. Por otro lado, el sistema de segmentación que usa el APEIM y el INEI, son para zonas urbanas, debido a que estos factores han sido elegidos, en función a la capacidad de segmentación y guía de desarrollo para políticas de estado y estudios de mercados para la venta, comercialización y distribución de productos y servicios de empresas privadas (Ver tabla 11).

TABLA XI
DIFERENCIAS ENTRE LA PROPUESTA DE CLASIFICACIÓN URBANA Y RURAL

	APEIM e INEI (Urbana)	Propuesta de rural
Educación	Si	No
Ocupación principal	No	Si
Tenencias de artefactos	Si	No
Material de la vivienda	Si	Si
Paredes	Si	Si
Techo	Si	Si
Piso	Si	Si
Consulta de salud	Si	Si
Tenencia de auto de uso particular y servicio doméstico (empleada del hogar)	Si	No
Tenencia de servicios públicos (internet, cable y telefonía fija)	Si	No
Tenencia de terrenos	No	Si
Dedicación al negocio	No	Si
Tipo de vivienda	No	Si

En la zona urbana, se deduce que la educación es un factor que mejora los ingresos, mientras que en la zona rural el nivel educativo no representa un factor de mejora en los ingresos económicos; de igual forma, la tenencia de artefactos. Finalmente, un factor clave para mejorar los niveles socioeconómicos en las zonas rurales, es la tenencia de terrenos en las zonas urbanas.

Entre sus limitaciones, se ha podido demostrar que se aplico netamente a un distrito rural, lo que origina una limitación en la base de datos, teniendo información sesgada hacia un lugar específico, pero su aplicación en el modelo respecto a la muestra, significa también la limitación respecto a los algoritmos que se puedan proponer, tales como el Random Forest, XGBoost o SVM.

VII. REFERENCIAS

[1] Asociación Peruana de Empresas de Investigación de Mercados (APEIM), Niveles socioeconómicos 2023–2024. Lima, Perú: APEIM, 2024. [En línea]. Disponible en: https://apeim.com.pe/wp-content/uploads/2024/01/APEIM-Informe-de-Niveles-Socioeconomicos-2023-2024-Version-WEB.pdf.

[2] V. Venkatesh, M. G. Morris, G. B. Davis y F. D. Davis, "User acceptance of information technology: Toward a unified view," MIS Quarterly, vol. 27, no. 3, pp. 425–478, 2003, doi: 10.2307/30036540.

[3] J. D. McLeod y E. R. Wright, Eds., The Sociology of Mental Illness: A Comprehensive Reader. New York, NY, USA: Oxford University Press, 2010.

[4] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006, doi: 10.1007/978-0-387-45528-0.

[5] Y. Bengio, A. Courville y P. Vincent, "Representation learning: A review and new perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, 2013, doi: 10.1109/TPAMI.2013.50.

[6] M. Ravallion y G. Datt, "How important to India's poor is the sectoral composition of economic growth?," World Bank Econ. Rev., vol. 10, no. 1, pp. 1–25, 1996, doi: 10.1093/wber/10.1.1.

[7] S. Faiguenbaum, C. Ortega y F. Soto Baquero, Eds., Pobreza rural y políticas públicas en América Latina y el Caribe, t. I. Santiago, Chile: FAO, 2013. [En línea]. Disponible en: https://www.fao.org/4/i3069s/i3069s.pdf.

[8] A. A. Manobanda Suárez et al., "Segmentación de mercados en tiempo real: Innovación y agilidad empresarial," Ciencia y Desarrollo, vol. 28, no. 1, pp. 541–550, 2025, doi: 10.21503/cyd.v28i1.2846.

[9] Centro Nacional de Planeamiento Estratégico (CEPLAN), Inteligencia artificial: Desafíos y oportunidades para el Perú. Lima, Perú: CEPLAN, 2021. [En línea]. Disponible en: https://www.gob.pe/institucion/ceplan/informes-publicaciones/2296616-inteligencia-artificial-desafios-y-oportunidades-para-el-peru.

[10] M. Ambrosio-Albalá y M. del M. Delgado-Serrano, "Understanding rural areas dynamics from a complex perspective: An application of prospective structural analysis," en Proc. 12th Congr. European Association of Agricultural Economists, Ghent, Belgium, 2008, doi: 10.22004/ag.econ.44159.

[11] R. Hernández-Sampieri y C. P. Mendoza Torres, Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta. Ciudad de México, México: McGraw-Hill Education, 2018.

[12] S. Fattah, A. R. Razak, M. A. Yusuf y A. A. R. Fernandes, "Econometric modelling of rural poverty, unemployment and the agricultural sector using a truncated spline approach with longitudinal data," Economies, vol. 13, no. 9, art. no. 273, 2025, doi: 10.3390/economies13090273.

[13] Y. Cardero Robert y J. Utria Romero, "Los criterios de segmentación y los métodos cuantitativos: Una exigencia para la elección de segmentos de mercado," Contribuciones a la Economía, vol. 10, no. 1, 2012. [En línea]. Disponible en: https://dialnet.unirioja.es/servlet/articulo?codigo=9174161.

[14] D. Fomo y A.-H. Sato, "Adaptive segmentation and statistical analysis for multivariate big data forecasting," Big Data Cogn. Comput., vol. 9, no. 11, art. no. 268, 2025, doi: 10.3390/bdcc9110268.

[15] S. O. Araújo, R. S. Peres, J. C. Ramalho, F. Lidon y J. Barata, "Machine learning applications in agriculture: Current trends, challenges, and future perspectives," Agronomy, vol. 13, no. 12, art. no. 2976, 2023, doi: 10.3390/agronomy13122976.

[16] L. K. David, J. Wang, W. Brooks y V. Angel, "Digital transformation and socio-economic development in emerging economies: A multinational analysis," Technology in Society, vol. 81, art. no. 102834, 2025, doi: 10.1016/j.techsoc.2025.102834.

[17] G. Avirappattu, A. Pach III, C. E. Locklear y A. Q. Briggs, "An optimized machine learning model for identifying socio-economic, demographic and health-related variables associated with low vaccination levels that vary across ZIP codes in California," Preventive Medicine Reports, vol. 28, art. no. 101858, 2022, doi: 10.1016/j.pmedr.2022.101858.

[18] R. Ramos Pollán, "Perspectivas y retos de las técnicas de inteligencia artificial en el ámbito de las ciencias sociales y de la comunicación," Anuario Electrónico de Estudios en Comunicación Social "Disertaciones", vol. 13, no. 1, pp. 21–34, 2020, doi: 10.12804/revistas.urosario.edu.co/disertaciones/a.7774.

[19] S. Park, “Global economic trends: Implications for international business strategy,” Academy of Accounting and Financial Studies Journal, vol. 28, no. 3, pp. 1–3, 2024. [En línea]. Disponible en: https://www.abacademies.org/articles/global-economic-trends-implications-for-international-business-strategy-16872.html.
[20] H. Saravia Poma, “Determinantes de la pobreza multidimensional en zonas rurales del Perú, 2023,” tesis de maestría, Facultad de Economía, Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú, 2026. [En línea]. Disponible en: https://hdl.handle.net/20.500.12773/21963.
[21] A. Schejtman y J. A. Berdegué, La desigualdad y la pobreza como desafíos para el desarrollo territorial rural, Documento de Trabajo no. 1. Santiago, Chile: Rimisp–Centro Latinoamericano para el Desarrollo Rural, 2007. [En línea]. Disponible en: https://www.rimisp.org/wp-content/files_mf/13590418601.pdf.
[22] S. Wang, B. Liu, Y. Yang, L. Yang y M. Zhen, “Urban-rural distinction or economic segmentation: A study on fear and inferiority in poor children’s peer relationships,” Healthcare, vol. 10, no. 10, art. no. 2057, 2022, doi: 10.3390/healthcare10102057.
[23] G. M. Rodríguez, “Desigualdad socioeconómica y segregación residencial en Argentina: Niveles y tendencias recientes (1991–2001–2010),” Cuadernos de Vivienda y Urbanismo, vol. 13, 2020, doi: 10.11144/Javeriana.cvu13.dssr.
[24] N. L. Ward, “Improving equity and access for low-income and minority youth into institutions of higher education,” Urban Education, vol. 41, no. 1, pp. 50–70, 2006, doi: 10.1177/0042085905282253.
[25] A. Çetinkaya, Ö. K. Baykan y H. Kırgız, “Analysis of machine learning classification approaches for predicting students’ programming aptitude,” Sustainability, vol. 15, no. 17, art. no. 12917, 2023, doi: 10.3390/su151712917.