

# Clasificación mediante Redes Neuronales y Conglomerados No Jerárquicos de las Condiciones de Vida de los Hogares de Bolivia

Lic. Luis Fernando Flores Rivero  
✉ luisfernandofloresr@gmail.com

## 1. Introducción

En los últimos años se han desarrollado metodologías para medir condiciones de vida de los hogares bolivianos, mediante indicadores de pobreza, tales como, el NBI, la incidencia de pobreza y la brecha de pobreza, que son indicadores descriptivos que resumen la información sin mostrar un panorama con la extensión de los datos como lo hace el análisis de conglomerados o las redes neuronales.

En este sentido, se pretende investigar cuál de los dos métodos (conglomerados no jerárquicos o redes neuronales) es óptimo.

Las redes neuronales artificiales son sistemas conexionistas dentro del campo de la inteligencia artificial, las cuales, dependiendo del tipo de arquitectura neuronal, pueden tener diferentes aplicaciones. Pueden utilizarse para el reconocimiento de patrones, la comprensión de información, el modelaje y la reducción de la dimensionalidad, el agrupamiento, la clasificación, etc.

El análisis de conglomerados reúne un conjunto de técnicas del análisis multivariante, cuya principal aplicación reside en la clasificación u obtención de tipologías que no están previamente identificadas.

## 2. Objetivo

El presente trabajo tiene como objetivo clasificar las condiciones de vida en los hogares en Bolivia con ambas metodologías, comparar sus resultados y analizar sus diferencias.

## 3. Metodología

Se utilizará al indicador de Necesidades Básicas Insatisfechas (NBI) como variable control o variable tipo. Dicho indicador servirá como punto de referencia para la evaluación de cada método. Además de lo anterior, el Indicador NBI permitirá evitar los problemas de variables con diferentes unidades o escalas de medida (proceso de estandarización).

### 3.1. Metodología NBI

El indicador NBI está conformado por cuatro componentes: Vivienda, Insumos y Servicios Básicos, Educación y Salud. Los pasos para el cálculo del indicador son los siguientes: 1) se determinan normas mínimas para cada uno de los componentes, 2) en relación a la norma, se elabora un índice de carencia que refleje la necesidad de un hogar para cada componente, 3) se procede a agregar todos estos componentes en un sólo índice de carencia (NBI) para cada hogar, 4) se agregan todos los hogares en cinco estratos de pobreza según su valor NBI (Figura 1).

### 3.2. Metodología de conglomerados

El análisis de conglomerados se divide en los grandes partes: El método jerárquico y el método no jerárquico. Por fines propios del trabajo se eligió el algoritmo no jerárquico de k-medias. Este procedimiento está dividido en cuatro etapas: 1) Establecer el número de conglomerados, 2) Seleccionar los centros iniciales, 3) Asignar los objetos a los grupos y 4) Optimizar o validar los resultados (Figura 2).

Figura 1 Metodología NBI

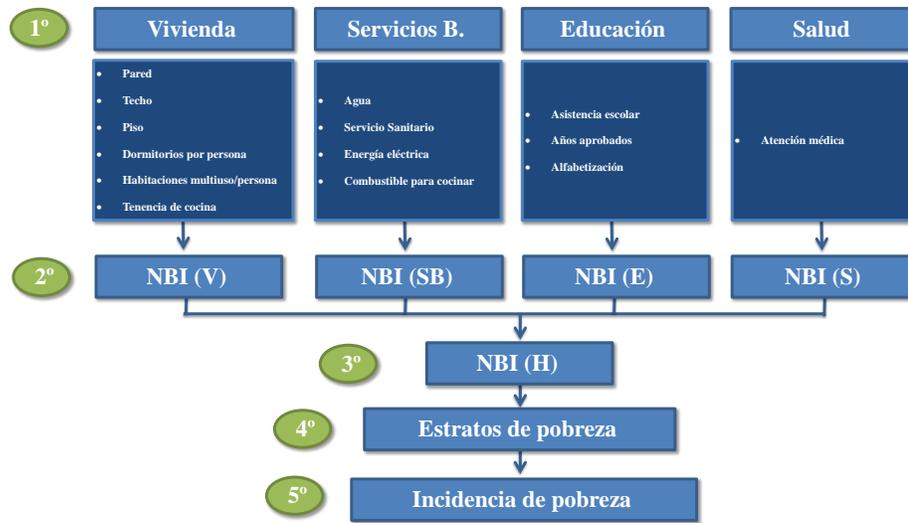


Figura 2 Metodología de conglomerados no jerárquicos



### 3.3. Metodología de la red neuronal

El modelo de red neuronal elegido para satisfacer las necesidades del trabajo, es la Red Neuronal Bacpropagation (BPN).

En una BPN existe una capa de entrada con “n” neuronas y una capa de salida con “m” neuronas y al menos una capa oculta de neuronas internas. Cada neurona de una capa (excepto las de entrada) recibe entradas de todas las neuronas de la capa anterior y envía su salida a todas las neuronas de la capa posterior (excepto las de salida). La estructura de la red BPN se muestra en la Figura 3.

La aplicación del algoritmo tiene dos fases, una hacia delante y otra hacia atrás. Durante la primera fase el patrón de entrada es presentado a la red y propagado a través de las capas hasta llegar a la capa de salida.

Obtenidos los valores de salida de la red, se inicia la segunda fase y se comparan los resultados con la salida esperada para así obtener el error. Se ajustan los pesos de la última capa proporcionalmente al error. Se pasa a la capa anterior con una retropropagación del error, ajustando los pesos y continuando con este proceso hasta llegar a la primera capa.

El proceso se repite hasta que el error sea nulo o se considere mínimo.

## 4. Análisis comparativo de resultados

Con el objetivo de verificar la precisión de cada método en la clasificación de las condiciones de vida de los hogares bolivianos, se compararon los resultados obtenidos de cada metodología y se analizaron las

posibles diferencias por estratos de pobreza.

Figura 3 ➤ Red Backpropagation

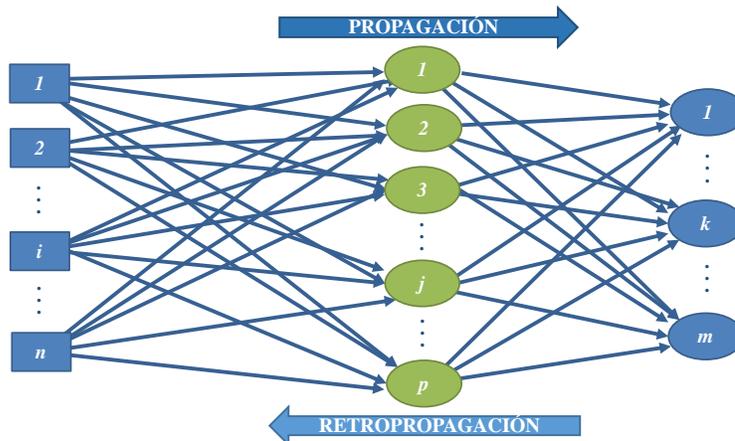


Figura 4 ➤ Metodología Red BPN

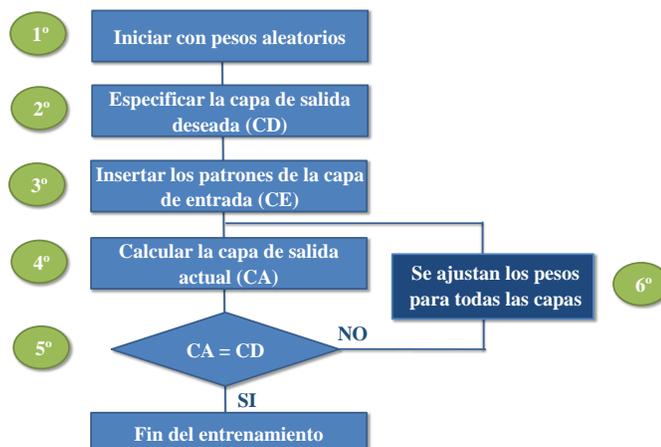
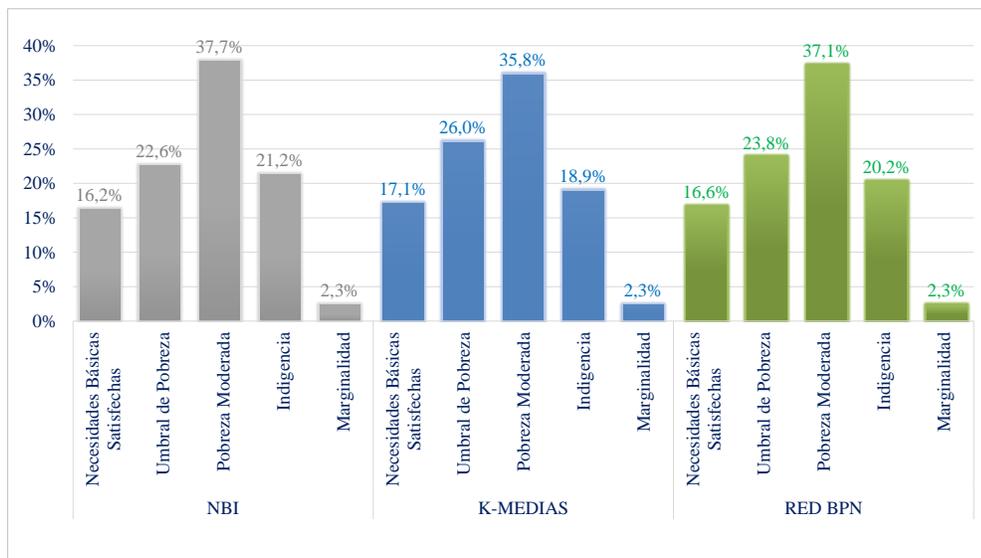


Figura 5 ➤ Estratos de pobreza según metodología de clasificación

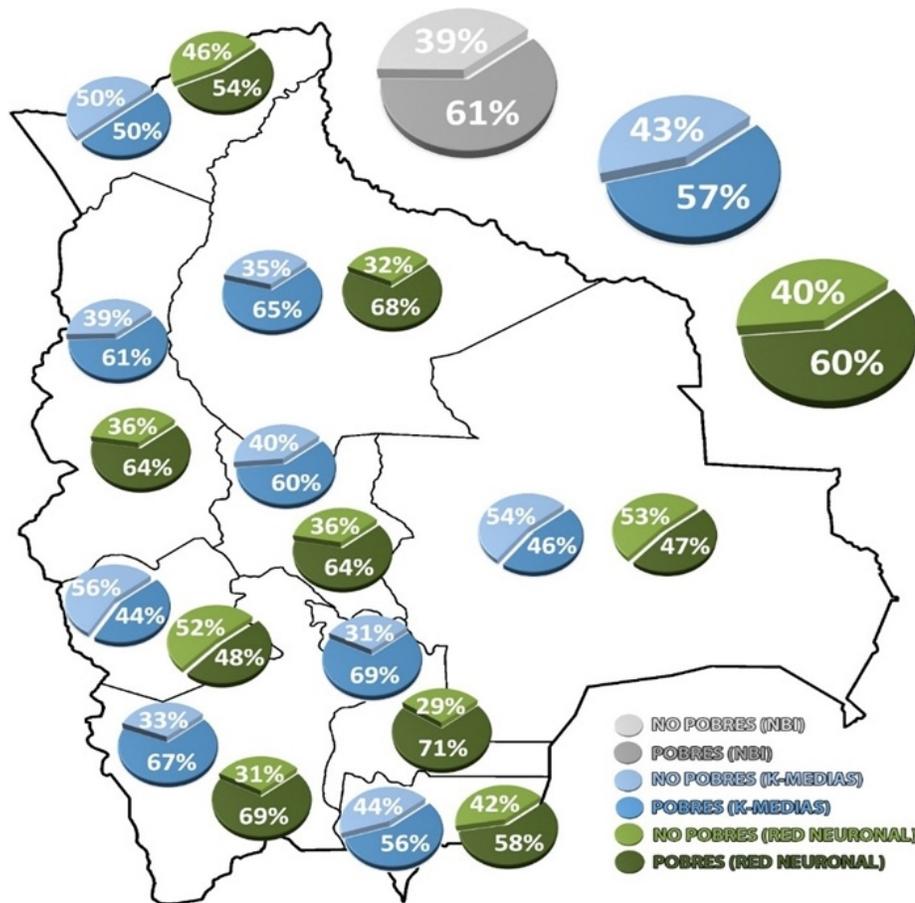


Luego, se contrastaron los resultados por condición de pobreza y a nivel departamental; y finalmente, se realizó el análisis de concordancia a través del estadístico Kappa, con el objetivo de verificar la optimalidad de cada método.

Según la Figura 5, la variable de control (NBI) muestra que la mayoría (37,7 %) de los hogares en Bolivia tienen pobreza moderada, seguido de un 22,6 % de hogares que cumplen con lo necesario para satisfacer sus necesidades básicas. Un 21,2 % está compuesto por hogares indigentes, tan solo el 16,2 % del total tiene satisfacción plena y el 2,3 % vive en condiciones marginales. Se observa también que los resultados obtenidos son bastante similares a los de la red neuronal BPN.

En relación a la variable control, el resultado de la clasificación por el método de conglomerados (k-medias) presenta menos hogares con pobreza moderada e indigencia, y aumenta considerablemente el umbral de pobreza en más de tres puntos porcentuales. Bajo esta clasificación, existirían menos hogares pobres en el territorio nacional.

Figura 6 ► Pobreza en los hogares de Bolivia según metodología de clasificación



Por otro lado y en referencia a la Figura<sup>1</sup> 6, se puede observar que según la clasificación NBI, el 61,0 % de los hogares en Bolivia son pobres y el 39,0 % no sufren de pobreza. Situación similar ocurre con los resultados de la red neuronal (40,0 % no pobres y 60,0 % pobres). Sin embargo la clasificación por conglomerados muestra mayor cantidad de hogares no pobres (43,0 %) y menor porcentaje de hogares pobres (57,0 %).

<sup>1</sup>Dado que el diseño muestral de la Encuesta de Hogares no incluye en su planteamiento la desagregación por departamento, se calcularon los errores de muestreo relativos para este nivel de desagregación y se obtuvieron resultados dentro del rango de aceptación.

Las mayores diferencias de clasificación se presentan en los departamentos de Cochabamba y Pando, seguidos de Oruro, La Paz, Beni y Tarija.

Según la metodología BPN, Chuquisaca es el departamento más pobre del país, con un 71,0 % de hogares que no cubren sus necesidades básicas; seguido de Potosí y Beni con niveles de pobreza del 69,0 % y 68,0 % respectivamente. Aproximadamente la mitad de los hogares del departamento de Oruro (48,0 %) son no pobres.

De acuerdo a las gráficas y al análisis anterior todo parece indicar que las redes neuronales logran clasificar de mejor forma las condiciones de vida de los hogares de Bolivia.

Con el fin de contrastar ambas metodologías y analizar sus diferencias se presenta la siguiente tabla:

**Tabla 1** Comparación de resultados por estrato de pobreza, según método de clasificación

MÉTODO	ESTRATOS DE POBREZA	Necesidades Básicas Satisfechas	Umbral de Pobreza	Pobreza Moderada	Indigencia	Marginalidad
<b>K-MEDIAS</b>	Necesidades Básicas Satisfechas	386.679	0	0	0	0
	Umbral de Pobreza	19.702	517.655	0	0	0
	Pobreza Moderada	0	101.006	796.977	0	0
	Indigencia	0	0	55.536	449.780	0
	Marginalidad	0	0	0	0	55.164
<b>TOTAL</b>		406.381	618.661	852.513	449.780	55.164
<b>RED BPN</b>	Necesidades Básicas Satisfechas	386.679	0	0	0	0
	Umbral de Pobreza	8.933	528.424	0	0	0
	Pobreza Moderada	0	37.716	860.267	0	0
	Indigencia	0	0	23.770	481.546	0
	Marginalidad	0	0	0	0	55.164
<b>TOTAL</b>		395.612	566.140	884.037	481.546	55.164

Fuente: Encuesta de Hogares. Elaboración propia.

Los valores de la diagonal principal, representan los casos clasificados de forma correcta y los valores fuera de la diagonal principal (resaltados en color celeste) representan los errores de clasificación.

**Tabla 2** Test de concordancia de kappa, según método de clasificación

Método	Valor	Significación aproximada	Número de casos válidos
K-Medias	0,90	0	2.382.499
Red BPN	0,96	0	2.382.499

Fuente: Encuesta de Hogares. Elaboración propia.

De acuerdo al estadístico Kappa, el método de conglomerados no jerárquicos alcanza un 90,0 % de objetos correctamente clasificados, mientras que la clasificación por redes neuronales muestra un 96,0 % de éxito. Por lo que se puede concluir que la técnica de clasificación por redes neuronales es 6 puntos porcentuales más efectiva.

La superioridad de las redes neuronales viene acompañada de recursos temporales y computacionales considerables. Por el contrario, y con menor nivel de precisión en la clasificación (10,0 %); la fácil accesibilidad, los pocos recursos informáticos y temporales, benefician el método por conglomerados no jerárquicos.

Finalmente y habiendo descrito brevemente los pros y contras de cada procedimiento, se deja a consideración del lector la elección de un método ante otro y se recomienda la utilización de ambos para un mayor

enriquecimiento y complementación en el análisis de resultados.

## Referencias

- [1] Choque Aspiazu, G., Loza Herrera, R. y Mendez Quintanilla, R. (2009). *Redes Neuronales Artificiales, Aplicaciones en MATLAB*. La Paz, Bolivia.
- [2] Freeman, J.A. and Skapura, D.M. (1993). *Redes Neuronales. Algoritmos, aplicaciones y técnicas de programación*. Addison Wesley/Díaz de Santos. USA.
- [3] Hecht-Nielsen, R. (1988). Neurocomputing: picking the human brain. *IEEE. Spectrum*, 25, No.3, 36-41.
- [4] Hiler, J. y Martínez, V. (1995). *Redes Neuronales Artificiales. Fundamentos, Modelos y aplicaciones*. Rama. Madrid, España.
- [5] Johnson, D.E. (2000). *Métodos multivariados aplicados al análisis de datos*. International Thomson Editores.
- [6] Luque Martínez, T. (2000). *Técnicas de análisis de datos en investigación de mercados*. Ediciones Pirámide. Madrid, España.
- [7] Martín, B. y Sanz, A. (2002). *Redes Neuronales y Sistemas Difusos*, 2ª ed. Alfaomega Rama. Madrid, España.
- [8] Peña, D. (2002). *Análisis de datos multivariantes*. McGRAW-HILL. Madrid, España.
- [9] Pérez, C. (2004). *Técnicas de Análisis Multivariante de Datos, Aplicaciones con SPSS*. Pearson Educación S. A. Madrid, España.