

INFOTEC CENTRO DE INVESTIGACIÓN E  
INNOVACIÓN EN TECNOLOGÍAS DE LA  
INFORMACIÓN Y COMUNICACIÓN

DIRECCIÓN ADJUNTA DE INNOVACIÓN Y  
CONOCIMIENTO  
GERENCIA DE CAPITAL HUMANO  
POSGRADOS

## “ÁREAS DE OPORTUNIDAD EN LA METODOLOGÍA DE MEDICIÓN DE LA POBREZA EN MÉXICO”

PROPUESTA DE INTERVENCIÓN  
Que para obtener el grado de MAESTRA EN  
CIENCIA DE DATOS E INFORMACIÓN

Presenta:  
**Areli Romero Cervantes**

Asesora:  
**Dra. Natalia García Colín**

Ciudad de México, junio, 2020.

**AUTORIZACIÓN DE IMPRESIÓN Y NO ADEUDO EN BIBLIOTECA**  
**MAESTRÍA EN CIENCIA DE DATOS E INFORMACIÓN**

Ciudad de México, 14 de octubre de 2020  
*INFOTEC-DAIC-GCH-SE-0564/2020.*

La Gerencia de Capital Humano / Gerencia de Investigación hacen constar que el trabajo de titulación intitulado

**ÁREAS DE OPORTUNIDAD EN LA METODOLOGÍA DE**  
**MEDICIÓN DE LA POBREZA EN MÉXICO**

Desarrollado por la alumna **Areli Romero Cervantes** y bajo la asesoría de la **Dra. Natalia García Colín**; cumple con el formato de biblioteca. Por lo cual, se expide la presente autorización para impresión del proyecto terminal al que se ha hecho mención:

Asimismo se hace constar que no debe material de la biblioteca de INFOTEC.

Vo. Bo.



**Mtra. Julieta Alcibar Hermosillo**  
Coordinadora de Biblioteca

**Anexar a la presente autorización al inicio de la versión impresa del trabajo referido que ampara la misma.**

## Agradecimientos

Agradezco a mis padres y hermanos por su apoyo incondicional siempre.

# Tabla de contenido

<b>Introducción.....</b>	<b>1</b>
<b>Capítulo 1. Objeto de la investigación.....</b>	<b>4</b>
1.1 Objetivo General .....	4
1.2 Objetivos específicos.....	4
1.3 Estado del arte .....	5
1.4 Hipótesis .....	8
<b>Capítulo 2. ¿Cómo clasifica la pobreza el CONEVAL? .....</b>	<b>10</b>
2.1 ¿Qué es el CONEVAL? .....	10
2.2 ¿Qué es la Encuesta Nacional de Ingresos y Gastos de los Hogares? .....	10
2.3 La metodología de medición de la pobreza .....	11
2.3.1. El espacio de bienestar.....	11
2.3.2. El espacio de derechos.....	13
2.4 Clasificación de la pobreza.....	14
2.5 Profundidad e intensidad de la pobreza.....	16
2.6 Sobre el conjunto de datos. ....	17
<b>Capítulo 3. Experimentos .....</b>	<b>21</b>
3.1 Parte 1. Análisis exploratorio .....	21
3.2 Parte 2. Una clasificación alternativa de la pobreza.....	30
3.2.1 Regresión logística multinomial.....	31
3.2.2 Naïve Bayes .....	35
3.2.3 Agrupamiento (Clustering) .....	40
3.2.4 Resultados comparados .....	48
<b>Conclusiones .....</b>	<b>52</b>
<b>Bibliografía.....</b>	<b>55</b>
<b>ANEXO 1 Referencias particulares sobre los algoritmos implementados ....</b>	<b>59</b>
<b>Índice de términos.....</b>	<b>61</b>

## Índice de figuras

<u>Figura 1. Población en situación de pobreza multidimensional.....</u>	<u>16</u>
--	-----------

## Índice de gráficos

<u>Gráfico 1. Distribución del ICTPC por condición de pobreza, 2016.....</u>	<u>25</u>
<u>Gráfico 2. Distribución del ICTPC por condición de pobreza, 2018.....</u>	<u>25</u>
<u>Gráfico 3. Gráfico de caja del ICTPC por condición de pobreza, 2016.....</u>	<u>27</u>
<u>Gráfico 4. Gráfico de caja del ICTPC por condición de pobreza, 2018.....</u>	<u>27</u>
<u>Gráfico 5. ICTPC promedio por decil de ingresos 2016 y 2018.....</u>	<u>28</u>
<u>Gráfico 6. Clasificación de pobreza de con la metodología del CONEVAL, 2016.....</u>	<u>29</u>
<u>Gráfico 7. Clasificación de pobreza de acuerdo con la metodología del CONEVAL, 2018.....</u>	<u>30</u>
<u>Gráfico 8. Clasificación de la condición de pobreza predicha por la regresión logística multinomial, 2016.....</u>	<u>32</u>
<u>Gráfico 9. Clasificación de la condición de pobreza predicha por la regresión logística multinomial, 2018.....</u>	<u>33</u>
<u>Gráfico 10. Clasificación de la condición de pobreza predicha por Naïve Bayes, 2016.....</u>	<u>27</u>
<u>Gráfico 11. Clasificación de la condición de pobreza predicha por Naïve Bayes, 2018.....</u>	<u>38</u>
<u>Gráfico 12. Clasificación de la condición de pobreza predicha con algoritmos de agrupamiento, 2016.....</u>	<u>42</u>
<u>Gráfico 13. Clasificación de la condición de pobreza predicha con algoritmos de agrupamiento, 2018.....</u>	<u>42</u>

Gráfico 14. Diferencias identificadas en los grupos por el contraste de carencias de derechos y el ingreso, 2016.....45

Gráfico 15. Diferencias identificadas en los grupos por el contraste de carencias de derechos y el ingreso, 2018.....47



## Índice de cuadros

<u>Cuadro 1. Construcción del ingreso corriente total.....</u>	<u>12</u>
<u>Cuadro 2. Componentes de los indicadores del espacio de bienestar.....</u>	<u>13</u>
<u>Cuadro 3. Descripción general del conjunto de datos que integra el CONEVAL... </u>	<u>18</u>
<u>Cuadro 4. Contraste de indicadores de carencia respecto la categoría de pobreza para 2016.....</u>	<u>22</u>
<u>Cuadro 5. Contraste de indicadores de carencia respecto la categoría de pobreza para 2018.....</u>	<u>23</u>
<u>Cuadro 6. Contraste de indicadores de carencia respecto el índice de privación, 2016.....</u>	<u>23</u>
<u>Cuadro 7. Contraste de indicadores de carencia respecto el índice de privación, 2018.....</u>	<u>24</u>
<u>Cuadro 8. Matriz de confusión resultados CONEVAL respecto regresión logística multinomial, 2016.....</u>	<u>33</u>
<u>Cuadro 9. Matriz de confusión resultados CONEVAL respecto regresión logística multinomial, 2018.....</u>	<u>34</u>
<u>Cuadro 10. Matriz de confusión resultados CONEVAL respecto Naïve Bayes, 2016.....</u>	<u>38</u>
<u>Cuadro 11. Matriz de confusión resultados CONEVAL respecto Naïve Bayes, 2018.....</u>	<u>39</u>
<u>Cuadro 12. Matriz de confusión resultados CONEVAL respecto algoritmos de agrupamiento, 2016.....</u>	<u>43</u>
<u>Cuadro 13. Matriz de confusión resultados CONEVAL respecto algoritmos de agrupamiento, 2018.....</u>	<u>43</u>
<u>Cuadro 14. Proporción de carencias respecto del total de población en cada clúster, 2016.....</u>	<u>44</u>

Cuadro 15. Proporción de carencias respecto del total de población en cada clúster, 2018.....46

Cuadro 16. Comparación de resultados con algoritmos de aprendizaje supervisado, 2016.....48

Cuadro 17. Comparación de resultados con algoritmos de aprendizaje supervisado, 2018.....49

Cuadro 18. Comparación de resultados aprendizaje no supervisado, 2016 y 2018.....50

## Siglas y abreviaturas

<b>CONEVAL</b>	Consejo Nacional de Evaluación de la Política de Desarrollo Social.
<b>ENIGH</b>	Encuesta Nacional de Ingresos y Gastos de los Hogares.
<b>ICTPC</b>	Ingreso corriente total per cápita.
<b>IPM</b>	Índice de Pobreza Multidimensional.
<b>INEGI</b>	Instituto Nacional de Estadística y Geografía.
<b>LGDS</b>	Ley General de Desarrollo Social.
<b>LPI</b>	Línea de pobreza por ingresos.

## Glosario

### “A”

**Agrupamiento (Clustering):** Algoritmos cuyo objetivo es segmentar los registros en grupos, de modo que las diferencias entre los asignados al mismo grupo tienden a ser más pequeñas que entre aquellos que se asignaron a grupos diferentes, en otras palabras, se agrupa a los más similares entre sí.

**Aprendizaje no supervisado:** Comprende algoritmos en los cuales no se cuenta con una variable objetivo y por lo tanto no se realiza un procedimiento de entrenamiento mediante dicha variable; sino que mediante la determinación de similitud entre los registros estos se asocian entre sí. Debido a la ausencia de la variable objetivo las validaciones de efectividad del algoritmo son más complejas.

**Aprendizaje supervisado:** Se entrena el algoritmo para que aprenda de la interacción entre los atributos y una variable objetivo y de este aprendizaje pueda determinar las clases a las que pertenece cada registro. En el aprendizaje supervisado es más sencillo identificar si las predicciones fueron exitosas o no debido a la existencia de la variable objetivo, ya que con ella se pueden contrastar los resultados.

### “C”

**Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL):** Es un organismo público descentralizado, con personalidad jurídica, patrimonio propio, autonomía técnica y de gestión. Tiene por objeto normar y coordinar la evaluación de las Políticas y Programas de Desarrollo Social, que ejecuten las dependencias públicas, y establecer los lineamientos y criterios para la definición, identificación y medición de la pobreza, garantizando la transparencia, objetividad y rigor técnico en dicha actividad.

## “D”

**Doble condicionante:** Es que de forma simultánea no se cuenten con ingresos suficientes para cubrir sus necesidades básicas, es decir ingresos inferiores a la LPI y presentar al menos una carencia de derechos básicos.

## “E”

**Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH):** Tiene como objetivo proporcionar un panorama estadístico del comportamiento de los ingresos y gastos de los hogares en cuanto a su monto, procedencia y distribución; adicionalmente, ofrece información sobre las características ocupacionales y sociodemográficas de los integrantes del hogar, así como las características de la infraestructura de la vivienda y el equipamiento del hogar.

**Espacio de bienestar:** Se emplea en un contexto para medir la pobreza monetaria, para lo cual se compara el ICTPC con el valor de dos canastas: una alimentaria y otra que incluye, además de los alimentos, todos los bienes y servicios de consumo habitual. Su finalidad es crear una medida tangible que permite hacer una aproximación para determinar si se cumple la condición de disponibilidad de recursos económicos.

**Espacio de derechos:** Se construye sobre la premisa de que toda persona debe tener acceso a una serie de recursos que son necesarios para garantizar la dignidad humana y que además se encuentran en la normatividad como garantías mínimas de derecho. En el caso del espacio de derechos se establece el umbral que separa a los carentes de los no carentes.

## “J”

**Ingreso corriente total per cápita (ICTPC):** Se construye, a partir de información de la ENIGH, como una variable continua que permite medir en términos monetarios el espacio de bienestar.

**Índice de privación:** El índice de privación consiste en la suma de los seis indicadores de carencia; en consecuencia, el valor de dicho índice puede ascender como máximo a 6.

**Intensidad de la pobreza:** Es el producto de la medida de incidencia de la pobreza y la proporción promedio de carencias sociales de la población pobre multidimensional (CONEVAL, 2014) (CONEVAL, 2019).

**Índice de Pobreza Multidimensional (IPM):** Con el objeto de poder realizar comparaciones de pobreza en el mundo se construyó este índice el cual mide privaciones multidimensionales agudas. Utiliza información de 10 indicadores con determinada ponderación de tal forma que el máximo nivel del índice sea 100%. Dichos indicadores se dividen en tres dimensiones: salud, educación y estándares de vida; se considera que las personas son pobres si experimentan privación en por lo menos un tercio de los indicadores.

## “K”

**K-means:** Es un algoritmo de aprendizaje no supervisados que se basa en seleccionar aleatoriamente los objetos iniciales para identificar los más cercanos a éstos y así determina los primeros grupos, después determina la media de los grupos previamente definidos y recalcula la similitud entre cada objeto y dicha media y reubica a cada objeto en el grupo cuya media es más cercana, este proceso se realiza iterativamente. La meta del proceso es hacer que la sumatoria de todas las distancias respecto de su media sea la mínima posible.

**K-modes:** Es un algoritmo de aprendizaje no supervisado que considera únicamente datos categóricos y presenta los siguientes cambios respecto de K-means: a) usa una medida de discrepancias entre datos categóricos; b) reemplaza el uso de la media por la moda; y c) usa la frecuencia para localizar la modas en cada grupo, la medida de similitud se determina por el número de discrepancias entre dos objetos categóricos, así el menor número de discrepancias significa mayor similitud entre los objetos.

**K-prototypes:** La medida de similitud para la parte que corresponde a los datos numéricos se define simplemente por la distancia euclidiana al cuadrado, y sigue la misma lógica iterativa de K-means.

## “L”

**Línea de pobreza por ingresos (LPI):** Constituye el valor monetario de una canasta de alimentos, bienes y servicios básicos (CONEVAL, 2019) (CONEVAL, s.f.).

**Línea de pobreza extrema por ingresos:** Valor monetario de una canasta alimentaria básica.

## “M”

**Matriz de confusión:** Actúa como una tabla de contingencia que contribuye a identificar los registros que fueron clasificados o predichos correctamente comparados con cada clase de los datos originales. Por lo cual contribuye como herramienta para identificar las coincidencias o diferencias de clasificación.

**Medidas de incidencia:** Es el porcentaje de la población que padece algún tipo de carencia económica o social.

**Metodología de medición multidimensional de la pobreza:** De acuerdo con el CONEVAL, el bienestar económico y el acceso a derechos básicos son las características que determinan la condición de pobreza o no pobreza en una persona. Así: para que una persona sea considerada pobre es necesario que sea carente tanto en el espacio de bienestar como en el de derechos.

## “N”

**Naïve Bayes:** Este algoritmo pertenece al grupo de algoritmos de aprendizaje supervisado. El principal supuesto es que las variables que servirán de entrada para entrenar al algoritmo son independientes entre ellas y se fundamenta en la probabilidad condicional para determinar el impacto de un evento en otro. La determinación de la clase de cada individuo se determinará a partir de la clase que obtenga la probabilidad máxima, considerando la probabilidad de cada característica en la clase y ponderado por la probabilidad de la clase

**No pobre multidimensional y no vulnerable:** Población cuyo ingreso es superior a la LPI y que no tiene carencia social alguna (CONEVAL, 2014) (CONEVAL, 2019).

## “P”

**Pobreza extrema:** Se presenta cuando una persona observa más de 3 carencias y el ingreso se encuentra por debajo de la línea de pobreza extrema por ingresos.

**Pobres multidimensionales:** Población con ingreso inferior al valor de la LPI y que padece al menos una carencia social.

**Pobreza multidimensional:** Una persona se encuentra en situación de pobreza cuando tiene al menos una carencia social (en los seis indicadores de rezago educativo, acceso a servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, servicios básicos en la vivienda y acceso a la



alimentación) y su ingreso es insuficiente para adquirir los bienes y servicios que requiere para satisfacer sus necesidades alimentarias y no alimentarias. La palabra multidimensional se refiere a que la metodología de medición de pobreza utiliza varias dimensiones o factores económicos y sociales en su concepción y definición (CONEVAL, s.f.).

**Probabilidad condicional:** Es la consideración de que la probabilidad de un evento podría aumentar o disminuir si se sabe que otro evento influye en el evento en estudio. Es decir, es la probabilidad del evento B dado el evento A.

**Profundidad de la pobreza:** Se reportan dos medidas de profundidad una asociada al espacio del bienestar y otra respecto al índice de privación social. En el caso del espacio de bienestar se trata de una medida que está dada por la distancia promedio del ingreso de la población con un ingreso inferior a la LPI, respecto a esta misma línea. Esta medida se presenta tanto para la población con carencias sociales como para aquella que no las tiene. En cuanto al índice de privación social, la profundidad de las carencias se reporta mediante el número y la proporción promedio de carencias sociales. Este indicador debe estimarse para los siguientes grupos: la población en pobreza multidimensional, la población en pobreza multidimensional extrema, la población con ingresos superiores a la LPI que tiene al menos una carencia, y la población con al menos una carencia social.

## “R”

**Regresión logística:** Forma parte del grupo de algoritmos de clasificación en aprendizaje supervisado, este algoritmo tiene como objetivo modelar la probabilidad posterior, es decir la probabilidad de que un registro se clasifique en determinada clase dado que existe evidencia de que otro evento sucede. El algoritmo toma como entrada el número K de clases y para cada registro asigna un valor de probabilidad entre 0 y 1, que indicará la posibilidad de que se clasifique en una clase o en otra. Cuando  $K=2$ , se dice que se tiene una variable objetivo binaria. En este caso el modelo es simple al tratarse de una función lineal

y puede ser muy útil para clasificaciones donde se busca determinar si cierta característica está presente o ausente.

**Regresión logística multinomial:** Se diferencia de la regresión logística por que  $K > 2$ , es decir se tiene una variable objetivo categórica con más de dos categorías y se crea un modelo de regresión lineal para cada clase y luego se combina en una sola ecuación asegurando que la restricción de probabilidad de la suma de todas las probabilidades sea 1.

“V”

**Vulnerables por carencias sociales:** Población que presenta una o más carencias sociales, pero cuyo ingreso es superior a la LPI (CONEVAL, 2014) (CONEVAL, 2019).

**Vulnerables por ingresos:** Población que no presenta carencias sociales y cuyo ingreso es inferior o igual a la LPI (CONEVAL, 2014) (CONEVAL, 2019).

## Introducción

En esta investigación se muestra que la metodología que implementa el CONEVAL<sup>1</sup> para la identificación y por lo tanto la medición de la pobreza presenta áreas de oportunidad, debido a que dicha metodología no identifica de forma amplia y profunda la totalidad de la pobreza. A partir de algoritmos de aprendizaje supervisado y no supervisado se obtienen clasificaciones alternativas de pobreza que permiten observar como la metodología del CONEVAL al ser muy estricta, ya que para que una persona sea considerada pobre debe cumplir con una doble condición, que implica que de forma simultánea no se cuenten con ingresos suficientes para cubrir sus necesidades básicas y presentar al menos una carencia de derechos básicos, sin embargo pueden dejarse al margen personas cuyas condiciones económicas y sociales son muy similares a las de hogares que son clasificados como pobres. Así, entre los resultados más importantes, se observa que existen hogares clasificados por el CONEVAL como vulnerables por carencias, cuyas condiciones de vida son equivalentes a las de personas clasificadas como pobres. De acuerdo con lo anterior, la actual metodología subestima los niveles de pobreza y con ello no únicamente se sesga la visión sobre la dimensión de la pobreza en México, sino también resulta desfavorable para el diseño de políticas públicas encaminadas a resolver el problema.

Una correcta medición de la pobreza permite dimensionar de forma adecuada esta problemática social contribuyendo a fortalecer la capacidad de afrontar dicho problema de manera eficaz. Por ésta razón, es importante robustecer la metodología con la cual se mide la pobreza en México. En esta tesis se presenta un análisis de la metodología que usa el CONEVAL para la medición de pobreza, y posteriormente se proponen metodologías que pueden brindar una

---

<sup>1</sup> De acuerdo con la LGDS en su artículo 81 el CONEVAL es un organismo público descentralizado, con personalidad jurídica, patrimonio propio, autonomía técnica y de gestión de conformidad con la Ley Federal de las Entidades Paraestatales. Tiene por objeto normar y coordinar la evaluación de las Políticas y Programas de Desarrollo Social, que ejecuten las dependencias públicas, y establecer los lineamientos y criterios para la definición, identificación y medición de la pobreza, garantizando la transparencia, objetividad y rigor técnico en dicha actividad.

perspectiva distinta y enriquecedora sobre éste fenómeno. Para ello se usa la información que en México genera el INEGI mediante la ENIGH<sup>2</sup> que es una encuesta que se aplica a hogares y que recopila información socioeconómica que permite al CONEVAL generar los datos necesarios para aplicar su metodología.

El presente estudio tiene como objeto, identificar a partir de los mismos atributos que usa el CONEVAL para su clasificación de pobreza y mediante la implementación de algoritmos de aprendizaje supervisado y no supervisado, patrones comunes entre la población que permitan clasificarla como pobre, vulnerable o no pobre, o bien agruparla con individuos cuyas características son más parecidas. Se espera localizar áreas de oportunidad en la metodología analítica del CONEVAL, con la finalidad de añadir herramientas que permitan examinar y atender mejor el problema.

En el Capítulo 1 se presenta y describe el objeto de la investigación. En el Capítulo 2 se describe de forma detallada la implementación de la metodología de clasificación que usa el CONEVAL y una descripción de la forma en la que el CONEVAL recopila y construye la base de datos mediante la cual se mide la pobreza. El Capítulo 3 se divide en dos partes; en la primera se presenta un análisis descriptivo de la base de datos utilizada por el CONEVAL; en la segunda parte se contrastan los resultados obtenidos por el CONEVAL con aquellos que obtenemos mediante el uso de algoritmos de clasificación de aprendizaje supervisado y de agrupamiento con aprendizaje no supervisado. Al finalizar los capítulos se presentan las conclusiones obtenidas de la investigación.

---

<sup>2</sup> La Encuesta Nacional de Ingresos y Gastos de los Hogares 2016 y 2018 se llevaron a cabo del 21 de agosto al 28 de noviembre de 2016 y 2018, respectivamente. Su objetivo es proporcionar un panorama estadístico del comportamiento de los ingresos y gastos de los hogares en cuanto a su monto, procedencia y distribución; adicionalmente, ofrece información sobre las características ocupacionales y sociodemográficas de los integrantes del hogar, así como las características de la infraestructura de la vivienda y el equipamiento del hogar. (INEGI, 2019) (INEGI, 2017)



# Capítulo 1

## Objeto de la investigación

## Capítulo 1. Objeto de la investigación

La pobreza es un problema complejo que limita el potencial de una sociedad para desarrollarse en el ámbito económico y humano, restringe a las sociedades porque pone en desventaja a la población que se encuentra en dicha situación.

Por lo tanto, es importante entenderla y tener una noción amplia de su dimensión para poder atenderla con las políticas públicas adecuadas. En este sentido en esta investigación se busca proponer formas alternativas de medirla y para concluir que la metodología actual puede ser mejorada.

### 1.1 Objetivo General

A partir de un análisis de las métricas actuales implementadas por el CONEVAL sobre la pobreza en México, y de su contraste con métodos alternativos de clasificación o agrupación se buscan localizar áreas de oportunidad, que permitan sustentar la necesidad de robustecer la identificación, medición y análisis de pobreza que actualmente implementa el CONEVAL.

### 1.2 Objetivos específicos

1. Realizar un análisis exploratorio de los resultados que presenta el CONEVAL, con la finalidad de identificar las principales diferencias y similitudes entre sus resultados y los nuestros.
2. Mediante la implementación de algoritmos de aprendizaje supervisado y no supervisado se buscan identificar clasificaciones distintas de pobreza y analizar sus resultados para detectar aportes adicionales a las clasificaciones de pobreza elaboradas por el CONEVAL.
3. Localizar áreas de oportunidad en la metodología del CONEVAL, de acuerdo a los contrastes y resultados obtenidos que permitan argumentar la conveniencia de robustecer esta metodología.

### 1.3 Estado del arte

La identificación de la pobreza siempre ha sido una tarea compleja tanto en términos conceptuales como en el ámbito de los datos. Es importante señalar que existe una amplia literatura que ha evolucionado respecto de la definición de pobreza y los métodos para identificarla, en este sentido se describen dos alternativas que se desprenden del ámbito internacional respecto de la detección de pobreza y la necesidad de ampliar la perspectiva en que ésta es identificada.

En lo que respecta a la necesidad de unificar metodologías para las medidas de pobreza en el mundo; en el año 2018, se publicó un Índice que mide la Pobreza Multidimensional (IPM), con el objeto de poder realizar comparaciones de pobreza en el mundo; éste mide privaciones multidimensionales agudas en 105 países, cubriendo el 77% de la población de la tierra. El IPM utiliza información de 10 indicadores<sup>3</sup> con determinada ponderación de tal forma que el máximo nivel del índice sea 100%. Dichos indicadores se dividen en tres dimensiones: salud, educación y estándares de vida; se considera que las personas son pobres si experimentan privación en por lo menos un tercio de los indicadores (PNUD, 2019).

Por otro lado, la medición de la pobreza se ha centrado en identificar una línea que permita separar a los pobres de los no pobres, esta línea se suele medir en términos monetarios y se compara con los ingresos de las personas. Sin embargo, definir de forma adecuada esta línea no es una tarea sencilla, los elementos que suelen integrarla son la canasta de alimentos y necesidades básicas, así como las necesidades que no son alimentarias. No obstante, la línea de pobreza de ingresos fomenta la impresión de que una prueba estadística es capaz de identificar quién está y quién no está sufriendo el impacto del fenómeno de la pobreza, pero la experiencia de la pobreza es multifacética, multidimensional y sujeta a volatilidad; también tiene dimensiones de tiempo y relatividad (Naciones Unidas, 2009).

---

<sup>3</sup> En la dimensión de salud se encuentran: nutrición y mortalidad infantil; en la dimensión de educación; años de escolaridad, asistencia escolar; en la dimensión de estándar de vida: combustible para cocinar, saneamiento, agua potable, electricidad, materiales de la vivienda, activos de la vivienda.

Así, la pobreza no puede definirse por una línea recta que divide una población en dos segmentos según un criterio de gasto o ingreso. Los déficits y privaciones en varias dimensiones específicas como educación, salud, vivienda, etc., comúnmente persisten en niveles de gasto muy por encima de lo estipulado a la línea de pobreza. Por lo tanto, se necesitan múltiples indicadores además de medidas de ingresos y gastos monetarios para capturar la escala y las dimensiones de la pobreza. Si bien la pobreza monetaria puede afectar solo a una minoría, solo unos pocos pueden escapar de la pobreza en cualquier otra forma. (Naciones Unidas, 2009).

La medición de pobreza tradicionalmente se realiza a partir de datos de encuestas en hogares, tal como se hace en México, posteriormente a dichos datos se les aplica la metodología utilizada por el CONEVAL para determinar la pobreza. Sin embargo, en muchos países en desarrollo, sobre todo tratándose de países del continente africano, se ha observado que éstas pueden ser limitadas o bien los datos no se consideran completamente confiables. En esos casos es necesario recurrir a otro tipo de datos que se recolectan sistemáticamente y que, aunque no necesariamente se recolectan con este fin lo cual conlleva a debilidades en la información, también pueden ser muy útiles. En este sentido, destacan tres estudios que usaron inteligencia artificial detectar pobreza y riqueza:

Investigadores de la Universidad de Stanford aplicaron técnicas de análisis de datos a millones de imágenes satelitales de alta resolución de países de cinco países de África<sup>4</sup>, para tener una aproximación del gasto de consumo de los hogares y su riqueza en activos. Los investigadores encontraron que, aunque las luces nocturnas muestran poca variación a niveles de gasto más bajos, los datos que se presentan en encuestas realizadas indican que otras características visibles en las imágenes satelitales diurnas, como el material del techo y la distancia con áreas urbanas, varían de forma correspondiente con el gasto y, por lo tanto, capturan mejor la variación entre los grupos más pobres. Las luces nocturnas tampoco distinguen entre zonas pobres densamente pobladas y ricas, o

---

<sup>4</sup> Nigeria, Tanzania, Uganda, Malawi y Ruanda



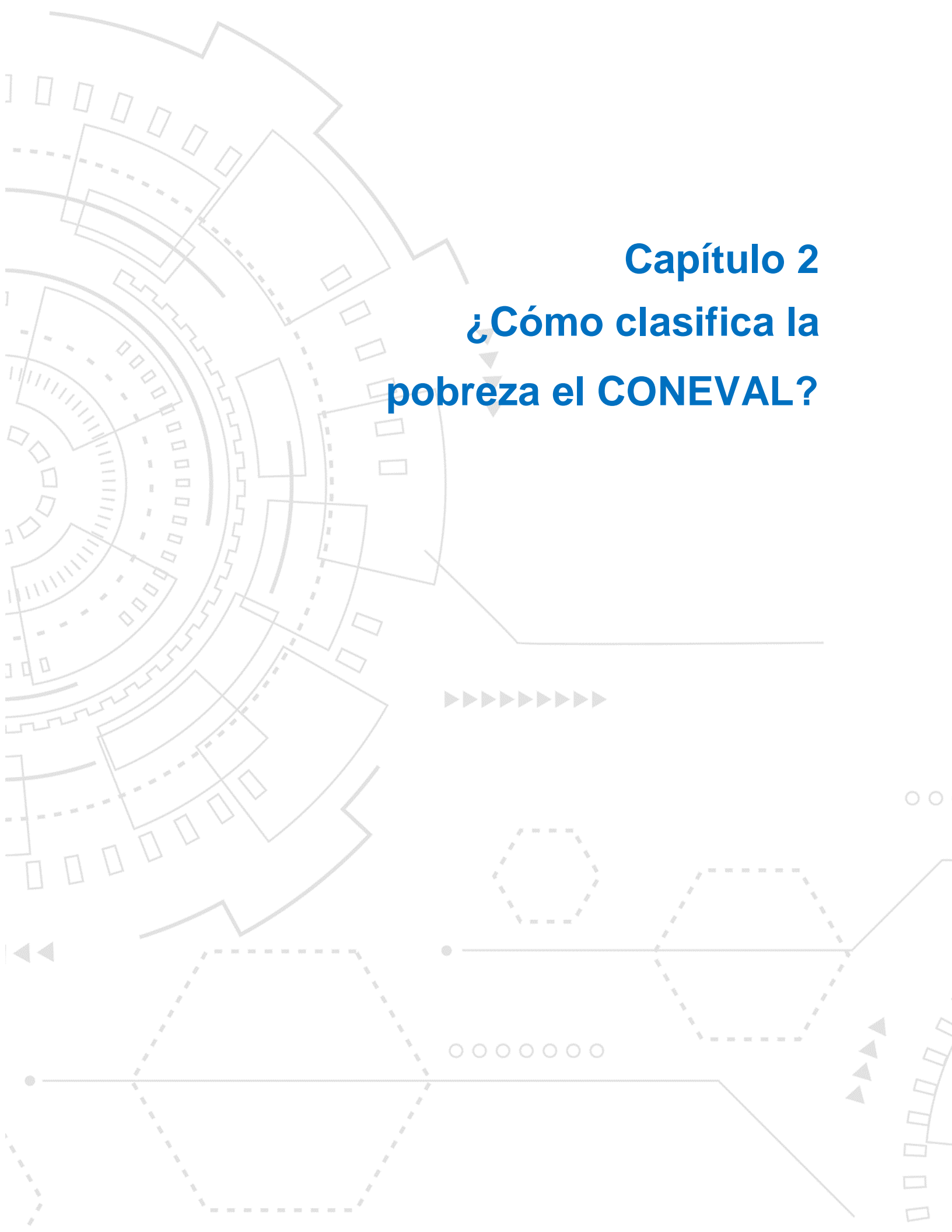
bien detectar esta misma diferencia entre zonas poco pobladas, pero los datos de luces nocturnas se usan como datos intermedios para aprender características de imagen que están correlacionadas con bienestar económico. En las imágenes tomadas en el día se aprecian factores como áreas urbanas, carreteras, cuerpos de agua, zonas rurales, entre otras; y aunque al modelo no se le solicita que busque tales características, ni que pudieran correlacionarse con la economía, el modelo aprende por sí mismo que estas características son útiles para estimar la intensidad de las luces en las noches. Así, la combinación de imágenes nocturnas y esta información de día hace que las predicciones del modelo tengan buena aproximación con los gastos promedio de consumo familiar y la riqueza de activos (Jean, y otros, 2016).

Otra investigación usa el historial de los datos de teléfonos móviles en Ruanda; tales como como la frecuencia y el tiempo de llamadas, las redes de contacto, patrones de viaje y localización, y el historial de consumo y gasto, y los somete a métodos de análisis que contribuyen a determinar el estatus de pobreza o riqueza de los usuarios. Además, también muestran que los atributos predichos de millones de individuos pueden, a su vez, reconstruir con precisión la distribución de la riqueza de una nación entera o inferir la distribución de activos de las microrregiones compuestas de solo unos pocos hogares (Blumenstock, Cadamuro, & Robert, 2015). Existe otra investigación más reciente cuyo enfoque se basa en usar múltiples fuentes de imágenes públicas del espacio que aportan mayor información, para inferir diferencias espaciales y temporales en el bienestar económico a nivel local en el África subsahariana, incluso para países donde aún no existen datos confiables de encuestas y donde los métodos de interpolación basados en encuestas podrían tener dificultades para generar estimaciones precisas. En este sentido el modelo predice en promedio el 70% de la variación de la riqueza basada en el suelo a través de los años. Así, no solo se logra identificar los países más ricos de África de los más pobres, sino también diferencia los niveles de riqueza. A partir de estos datos se puede tener una perspectiva sobre los cambios de la riqueza a través del tiempo (Burke, y otros, 2020).

Las medidas de la pobreza son materia de constante estudio y evolución y se pueden observar estudios donde se proponen alternativas para la definición de pobreza tanto en términos teóricos como técnicos y por otro lado existen tendencias novedosas de estudios que se basan en la aplicación de modelos que se nutren de datos que se obtienen principalmente de imágenes satelitales o de historiales de teléfonos móviles, debido a que la información de las encuestas es limitada o no confiable; así la diferencia principal respecto del presente estudio es que los datos sí se toman de la encuesta realizada en hogares, es decir la ENIGH, y con estos datos se busca entrenar algoritmos de clasificación y agrupación para detectar la presencia o ausencia de pobreza en la población.

#### **1.4 Hipótesis**

En la construcción de la metodología para la medición de la pobreza multidimensional, el CONEVAL dice que una persona se considera pobre si presenta al menos una carencia de derechos y no cuenta con ingresos monetarios por encima de la LPI, es decir, de acuerdo con dicha metodología para que una persona sea clasificada como pobre es necesario que se cumplan las dos condiciones de deficiencia tanto en el espacio de bienestar como en el de derechos. Se considera que esta doble condicionante es estricta para la clasificación de pobreza y excluye de esta categoría a personas actualmente consideradas vulnerables por ingresos o carencias que enfrentan condiciones adversas. Así, usando las mismas variables **se busca contrastar la clasificación que realiza el CONEVAL con algoritmos de clasificación y agrupamiento, esperando localizar diferencias y áreas de oportunidad en la actual clasificación de pobreza.**



## Capítulo 2

# ¿Cómo clasifica la pobreza el CONEVAL?

## **Capítulo 2. ¿Cómo clasifica la pobreza el CONEVAL?**

### **2.1 ¿Qué es el CONEVAL?**

El CONEVAL es el organismo del estado mexicano encargado de establecer los lineamientos y criterios para la medición de la pobreza donde se deben cumplir dos objetivos principales: cumplir con la LGDS e identificar la población en situación de pobreza, lo anterior con la finalidad de implementar políticas públicas que permitan hacer frente al problema de la pobreza en México.

Para dar cumplimiento con la normatividad correspondiente el CONEVAL usa la información de la ENIGH cuya publicación es bianual, así, las mediciones de pobreza únicamente pueden hacerse con la misma periodicidad.

### **2.2 ¿Qué es la Encuesta Nacional de Ingresos y Gastos de los Hogares?**

La ENIGH es diseñada y aplicada por el INEGI y su objetivo es proporcionar un panorama estadístico de comportamiento de los ingresos y gastos de los hogares, adicionalmente permite recabar información sobre características socio-demográficas de la población que integra los hogares y sobre la infraestructura de servicios públicos a la que tienen acceso los miembros de los hogares, así como las condiciones de la vivienda de éstos.

Dicho instrumento es útil porque recopila información necesaria para identificar la situación de la población en materia de bienestar (como la determinación del ICTPC) y también en lo que respecta a derechos (educación, salud, seguridad social, alimentación, vivienda y sus servicios).

La ENIGH se realiza mediante la aplicación de cuestionarios a una muestra de hogares seleccionados, de tal forma que los resultados permitan inferir con confiabilidad estadística la cuantificación, comportamiento de los ingresos, gastos y otras características socioeconómicas del universo de hogares del país.

La información se recopilada en un modelo relacional de tablas y se hace pública en la página electrónica del INEGI. Usando la información de la ENIGH, el CONEVAL genera una tabla donde las filas representan a cada individuo y las columnas corresponden a las variables recopiladas, así como de variables construidas a partir de dicha información asociadas principalmente a indicadores de ingresos y carencias.

## **2.3 La metodología de medición de la pobreza**

El CONEVAL desarrolló la metodología para la medición multidimensional de la pobreza en México con base en la Ley General de Desarrollo Social (LGDS) (CONEVAL, 2014) (CONEVAL, 2019).

**Así, de acuerdo con el CONEVAL, el bienestar económico y el acceso a derechos básicos son las características que determinan la condición de pobreza o no pobreza en una persona.** De acuerdo con la metodología del CONEVAL para que una persona sea considerada pobre es necesario que sea carente tanto en el espacio de bienestar como en el de derechos (CONEVAL, 2014) (CONEVAL, 2019).

### **2.3.1. El espacio de bienestar**

En el caso concreto de la medición del espacio de bienestar, se emplea el método estándar de pobreza monetaria, para lo cual se compara el ICTPC con el valor de dos canastas: una alimentaria y otra que incluye, además de los alimentos, todos los bienes y servicios de consumo habitual. Es importante mencionar que, en el contexto del espacio de bienestar, el principal objetivo consiste en identificar las dimensiones y condiciones que limitan la libertad de las personas para desarrollarse plenamente. Y la finalidad es crear una medida tangible que permite hacer una aproximación para determinar si se cumple la condición de disponibilidad de recursos económicos (CONEVAL, 2014) (CONEVAL, 2019).

Para la medición del espacio de bienestar, el CONEVAL crea una variable de ICTPC la cual se construye, a partir de información de la ENIGH, como una variable continua que permite medir en términos monetarios el espacio de bienestar comparado con la LPI<sup>5</sup>. Así, la construcción del indicador del espacio de bienestar contempla los siguientes elementos (CONEVAL, 2014) (CONEVAL, 2019):

- Considera aquellos flujos monetarios y no monetarios que no pongan en riesgo o disminuyan los acervos de los hogares.
- Toma en cuenta la frecuencia de las transferencias y eliminar aquellas que no sean recurrentes.
- No incluye como parte del ingreso la estimación del alquiler o renta imputada.
- Considera las economías de escala y las escalas de equivalencia dentro de los hogares<sup>6</sup>.

	Ingreso corriente monetario	Ingreso corriente no monetario
<b>Ingreso Corriente Total</b>	1. Remuneraciones por trabajo subordinado. 2. Ingreso por trabajo independiente (incluye el autoconsumo). 3. Ingreso por renta de la propiedad. 4. Otros ingresos provenientes del trabajo. 5. Transferencias.	1. Pago en especie. 2. Transferencias en especie (regalos en especie, se excluyen las transferencias que se dan por única vez). 3. No incluye renta imputada.

**Cuadro 1.** Construcción del ingreso corriente total.  
Fuente: CONEVAL.

$$ICTPC = \frac{\text{ingresocorrientetotaldelhogar}}{\sum d_i n_i}$$

$n_i$ =el número de miembros del hogar en cada rango de edad  $i$ ;

<sup>5</sup> En la segunda edición de la Metodología para la medición multidimensional de la pobreza en México se denominaba Línea de Bienestar (CONEVAL, 2019) y constituye el valor monetario de una canasta de alimentos, bienes y servicios básicos (CONEVAL, s.f.).

<sup>6</sup> Las escalas de equivalencia ajustan el ingreso del hogar de acuerdo con el tamaño y las características demográficas de la familia (Rojas, 2020).

$d_i$ =es la escala de equivalencia (con economías de escala) que corresponde a cada grupo de edad.

### 2.3.2. El espacio de derechos

En el caso del espacio de derechos se establece el umbral que separa a los carentes de los no carentes. Así, de acuerdo con la LGDS se deben seguir tres decisiones metodológicas (CONEVAL, 2014) (CONEVAL, 2019):

- a. Las carencias deben medirse por variables dicotómicas; esto quiere decir que el derecho se satisface o no; al no haber término medio, la satisfacción no se gradúa en una escala ordinal;
- b. No hay un derecho superior a otro, por lo que todos deben tener el mismo peso;
- c. La carencia de cualquiera de los derechos hace que una persona sea carente.

En este sentido, el espacio de derechos se construye sobre la premisa de que toda persona debe tener acceso a una serie de recursos que son necesarios para garantizar la dignidad humana y que además se encuentran en la normatividad como garantías mínimas de derecho. Así, la condición de pobreza surge del no cumplimiento de dichas garantías (CONEVAL, 2014) (CONEVAL, 2019).

El espacio de derechos se mide a partir de seis indicadores de tipo binario, diseñados y contruidos con información capturada a través de la ENIGH y que permiten determinar si un individuo presenta la carencia de un derecho determinado. Los indicadores y la composición de cada uno de ellos se presentan en el Cuadro 2.

Indicador	Componentes
1. Indicador de carencia por rezago educativo.	Edad.
	Año de nacimiento.
	Inasistencia a la escuela.
	Nivel educativo.

<b>Indicador</b>	<b>Componentes</b>
2. Indicador de carencia por acceso a los servicios de salud.	Seguro Popular.
	IMSS.
	ISSSTE.
	Pemex, Defensa Nacional o Marina.
	Otros servicios médicos.
3. Indicador de carencia por acceso a la seguridad social.	Población económicamente activa.
	Tipo de trabajo.
	Servicios médicos.
	Afore.
	Servicios médicos voluntarios.
	Afore voluntaria.
	Jubilados y pensionados.
	Por condición de parentesco.
Por contar con el programa para adultos mayores.	
4. Indicador de carencia por calidad y espacios de la vivienda.	Indicador de carencia del material de piso de la vivienda.
	Indicador de carencia del material de techos de la vivienda.
	Indicador de carencia del material de muros de la vivienda.
	Indicador de carencia por índice de hacinamiento de la vivienda.
5. Indicador de carencia por acceso a los servicios básicos en la vivienda.	Indicador de carencia de acceso al agua en la vivienda.
	Indicador de carencia de servicio de drenaje en la vivienda.
	Indicador de carencia de servicios de electricidad en la vivienda.
	Indicador de carencia de servicio de combustible para cocinar en la vivienda
6. Indicador de carencia por acceso a la alimentación.	Total de inseguridad alimentaria en mayores de 18 años.
	Total de inseguridad alimentaria en menores de 18 años.

**Cuadro 2.** *Componentes de los indicadores del espacio de bienestar.*  
Fuente: Elaboración propia con información del CONEVAL.

## 2.4 Clasificación de la pobreza

Cuando se detecta que un individuo presenta por lo menos uno de los indicadores de carencia anteriormente mencionados y adicionalmente sus ingresos son



insuficientes para adquirir los bienes y servicios que requiere para satisfacer sus necesidades (es decir, cuyo ingreso es inferior a la LPI); se considera que dicho individuo se encuentra dentro de una condición de pobreza (CONEVAL, 2014) (CONEVAL, 2019).

Así, se considera que una persona experimenta carencias en el espacio de derechos sociales cuando el valor del índice de privación social es mayor a cero es decir ( $C=1$ ). El índice de privación consiste en la suma de los seis indicadores de carencia; en consecuencia, el valor de dicho índice puede ascender como máximo a 6. Por otro lado, el espacio de bienestar económico, el cual se mide por medio del ingreso de las personas, se separa en dos partes delimitadas por la LPI que permite diferenciar si las personas tienen ingreso suficiente o no (CONEVAL, 2014) (CONEVAL, 2019).

En este sentido, en la Figura 1 se identifica la clasificación de la población de acuerdo a su situación de pobreza multidimensional<sup>7</sup> o su ausencia de ella. Una vez determinado su ingreso y su índice de privación social, cualquier persona puede ser clasificada en uno, y sólo uno, de los siguientes cuadrantes (CONEVAL, 2014) (CONEVAL, 2019):

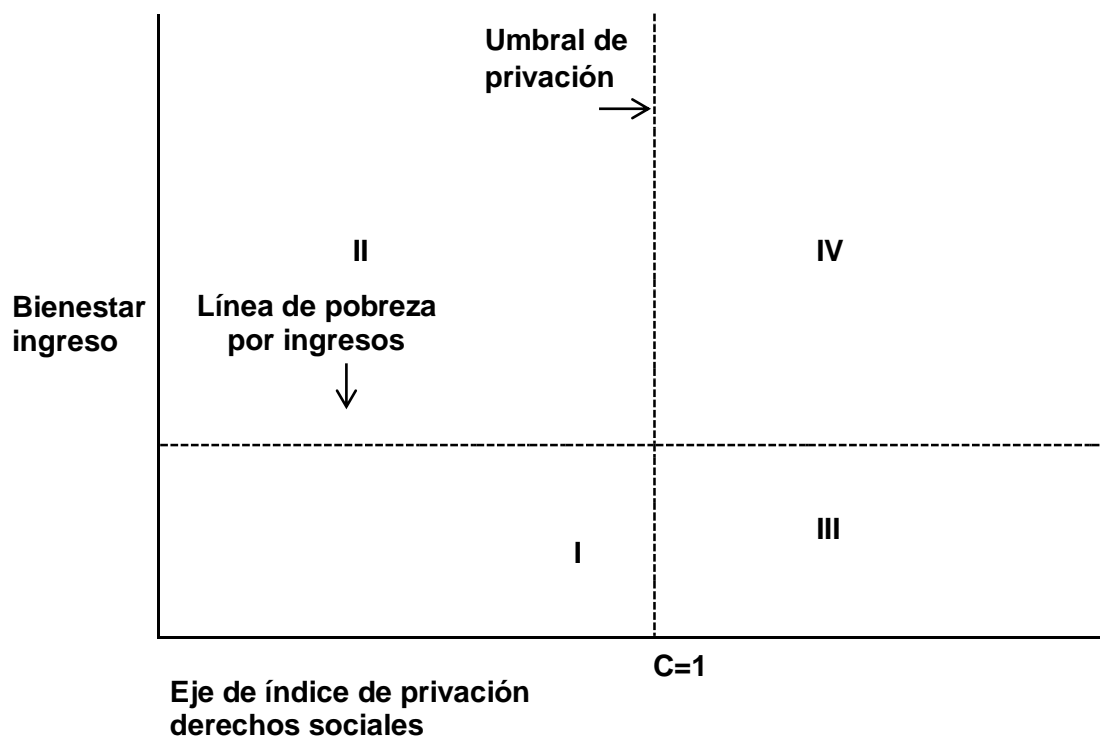
- I. Pobres multidimensionales: población con ingreso inferior al valor de la LPI<sup>8</sup> y que padece al menos una carencia social.
- II. Vulnerables por carencias sociales: población que presenta una o más carencias sociales, pero cuyo ingreso es superior a la LPI.
- III. Vulnerables por ingresos: población que no presenta carencias sociales y cuyo ingreso es inferior o igual a la LPI.

---

<sup>7</sup> Una persona se encuentra en situación de pobreza cuando tiene al menos una carencia social (en los seis indicadores de rezago educativo, acceso a servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, servicios básicos en la vivienda y acceso a la alimentación) y su ingreso es insuficiente para adquirir los bienes y servicios que requiere para satisfacer sus necesidades alimentarias y no alimentarias. La palabra multidimensional se refiere a que la metodología de medición de pobreza utiliza varias dimensiones o factores económicos y sociales en su concepción y definición (CONEVAL, s.f.).

<sup>8</sup> Valor total de la canasta alimentaria y de la canasta no alimentaria por persona, al mes. Donde la canasta alimentaria es el conjunto de alimentos cuyo valor sirve para construir la línea de pobreza extrema por ingresos. Éstos se determinan de acuerdo con el patrón de consumo de un grupo de personas que pueden satisfacer con ellos sus requerimientos de energía y nutrientes. La canasta no alimentaria se constituye de los bienes y servicios no alimentarios (CONEVAL, 2014) (CONEVAL, 2019) (CONEVAL, s.f.).

IV. No pobre multidimensional y no vulnerable: población cuyo ingreso es superior a la LPI y que no tiene carencia social alguna.



**Figura 1.** Población en situación de pobreza multidimensional.  
Fuente: CONEVAL.

## 2.5 Profundidad e intensidad de la pobreza.

Además de las medidas de identificación de pobreza mencionadas, se construyen otras métricas que permiten analizar la magnitud y la evolución de esta problemática. Estos criterios permiten agregar e integrar todas las características de los individuos y ampliar el concepto de pobreza multidimensional (CONEVAL, 2014) (CONEVAL, 2019):

- Medidas de incidencia: es el porcentaje de la población que padece algún tipo de carencia económica o social (CONEVAL, 2014) (CONEVAL, 2019).
- Profundidad: se reportan dos medidas de profundidad una asociada al espacio del bienestar y otra respecto al índice de privación social. En el caso del espacio de bienestar se trata de una medida que está dada por la distancia promedio del ingreso de la población con un ingreso inferior a la

LPI, respecto a esta misma línea. Esta medida se presenta tanto para la población con carencias sociales como para aquella que no las tiene. En cuanto al índice de privación social, la profundidad de las carencias se reporta mediante el número y la proporción promedio de carencias sociales. Este indicador debe estimarse para los siguientes grupos: la población en pobreza multidimensional, la población en pobreza multidimensional extrema, la población con ingresos superiores a la LPI que tiene al menos una carencia, y la población con al menos una carencia social (CONEVAL, 2014) (CONEVAL, 2019).

- Intensidad:
  - la intensidad de la pobreza multidimensional se define como el producto de la medida de incidencia de la pobreza y la proporción promedio de carencias sociales de la población pobre multidimensional (CONEVAL, 2014) (CONEVAL, 2019).
  - la intensidad de la pobreza multidimensional extrema se define como el producto de la incidencia de la pobreza multidimensional extrema y la proporción promedio de carencias sociales de la población pobre multidimensional extrema (CONEVAL, 2014) (CONEVAL, 2019).
  - la intensidad de la privación de la población con al menos una carencia se define como el producto de la medida de incidencia de la población que tiene al menos una carencia social y la proporción promedio de características de esa población (CONEVAL, 2014) (CONEVAL, 2019).

## **2.6 Sobre el conjunto de datos.**

A partir de la información que se obtiene de la ENIGH, el CONEVAL integra una base final que consta de 78 variables las cuales se pueden agrupar tal como se describe en el Cuadro 3.

Es importante señalar que de la ENIGH en la base del CONEVAL se mantiene la variable factor que ayuda a expandir los datos de la encuesta a nivel

de toda la población, Una de las características distintivas de las encuestas probabilísticas en hogares es que las viviendas seleccionadas y las personas que residen en ellas representan a otras viviendas y personas de condiciones socioeconómicas similares. El factor de expansión constituye el peso que se le da a cada unidad muestral para generalizar los resultados de la muestra a la población (INEGI, 2019).

Grupo	Tipo de Valor
Indicadores de carencia de derechos.	Variables categóricas binarias, donde 1 identifica la presencia de determinado indicador de carencia y 0 su ausencia.
Variables individuales que se extraen directamente de la ENIGH.	3 variables identifican la vivienda, el hogar y la persona. Edad y factor: variables numéricas discretas. niv_ed (Nivel educativo): variable categórica ordinal. sexo: variable binaria. tam_loc (Tamaño de la localidad): variable categórica nominal.
Variables auxiliares creadas específicamente para la construcción de los indicadores.	49 variables binarias, donde 1 es la presencia de la característica y 0 la ausencia. 4 variables categóricas nominales. 5 variables categóricas ordinales. 4 variables numéricas discretas.
Variable Ingreso corriente total per cápita e ingreso corriente total.	Variables numéricas continuas.

**Cuadro 3.** Descripción general del conjunto de datos que integra el CONEVAL<sup>9</sup>.  
Fuente: Elaboración propia con información del CONEVAL.

<sup>9</sup> Se debe señalar que la información descrita en el Cuadro 3. corresponde a la base final que crea el CONEVAL con información de la ENIGH 2016, para 2018 se observa una estructura similar pero el número de variables creadas es distinto.

Para realizar los análisis proyectados de este universo de datos se utilizaron los correspondientes a las publicaciones de 2016 y 2018, con el fin de observar cómo se comportan los resultados entre publicaciones. Y, con el fin de hacer comparaciones directas con la metodología del CONEVAL y observar las diferencias en las clasificaciones, se usaron para la implementación los 6 indicadores del espacio de derechos (educación, salud, seguridad social, alimentación, vivienda y sus servicios) que corresponden a variables binarias y uno del espacio de bienestar (ICTPC) que es una variable continua.

Se considera que las variables que se usaron aportan información esencial sobre las características mínimas tanto económicas como de derechos a las cuales una persona debe tener acceso para poder desarrollar todas sus aptitudes y talentos de tal forma que pueda acceder a un nivel de vida de calidad.

The background features a complex arrangement of light gray geometric elements. On the left, there are several overlapping circular and semi-circular shapes with various internal patterns, including dashed lines, small rectangles, and concentric circles. On the right, there are several horizontal and diagonal lines, some ending in arrowheads. A prominent feature is a dashed line forming a zigzag or hexagonal pattern. At the bottom right, there is a series of small, tilted rectangles arranged in a curve. The overall aesthetic is clean, technical, and modern.

# Capítulo 3

## Experimentos

## Capítulo 3. Experimentos

Como ya se mencionó con anterioridad, el CONEVAL construye los 6 indicadores de carencias (correspondientes al espacio de derechos) y el ICTPC (que alude al espacio de bienestar) y a partir de ellos determina la condición de pobreza de una persona. Los análisis que a continuación se presentan usaron los datos que se incorporan en las tablas finales que publica el CONEVAL, a partir de las cuales se realizaron las métricas de pobreza y cuyo origen es la ENIGH con información de 2016 y 2018.

Primero se realiza un análisis exploratorio del comportamiento de los indicadores de carencias y del ingreso, así como su interacción entre sí. Posteriormente, se usan estas variables como insumo para la implementación de algoritmos que permitan hacer clasificaciones de los registros y de esta forma contrastar la clasificación de pobreza hecha por CONEVAL y las clasificaciones o agrupamientos que realizan algoritmos de aprendizaje supervisado y no supervisado.

### 3.1 Parte 1. Análisis exploratorio

El CONEVAL considera que para ser clasificado en condición de pobreza se debe tener por lo menos una carencia en el espacio de derechos y no contar con ingresos superiores a la LPI, de lo cual se infiere que las personas que no cumplen con esta doble condición podrían presentar una u otra y no ser clasificados como pobres, en tal caso el CONEVAL les da la característica de vulnerables por ingresos o por carencias.

Con el objetivo de entender si la clasificación de la pobreza propuesta por el CONEVAL identifica las diferencias sustanciales entre las personas con distintas condiciones de pobreza, se presenta un análisis de las condiciones de carencia de derechos y de las distribuciones del ICTPC que se observaron en 2016 y 2018 para comprender los resultados de la actual clasificación de pobreza y sus áreas de oportunidad.

En los cuadros 4 y 5 se muestran respecto de la ausencia y presencia de todas las condiciones de pobreza, la estructura porcentual de los indicadores de carencia, para 2016 y 2018, respectivamente. Se destaca que el derecho a la seguridad social es la carencia de derechos observada con mayor proporción en población con pobreza para 2016 el 83% (de un total de 53,418,151 personas) presenta esta carencia y de la población vulnerable por carencias 74% (de un total de 32,860,754) tiene esta condición. En términos generales esta carencia se presenta en el 56% de la población de 2016.

Categoría de Pobreza	Población	Indicador de carencia por rezago educativo		Indicador de carencia por acceso a servicios de salud		Indicador de carencia por acceso a la seguridad social		Indicador de carencia por calidad y espacios de la vivienda		Indicador de carencia de acceso a servicios básicos en la vivienda		Indicador de carencia por acceso a la alimentación		
		A	P	A	P	A	P	A	P	A	P	A	P	
Pobreza	A	69,218,543	90%	10%	87%	13%	65%	35%	96%	4%	91%	9%	90%	10%
	P	53,418,151	73%	<b>27%</b>	81%	<b>19%</b>	17%	<b>83%</b>	78%	<b>22%</b>	67%	<b>33%</b>	67%	<b>33%</b>
Vulnerable por carencias	A	89,775,940	84%	16%	88%	12%	51%	49%	87%	13%	80%	20%	81%	19%
	P	32,860,754	79%	<b>21%</b>	73%	<b>27%</b>	26%	<b>74%</b>	91%	<b>9%</b>	82%	<b>18%</b>	78%	<b>22%</b>
Vulnerable por ingresos	A	114,031,672	81%	19%	83%	17%	40%	60%	87%	13%	79%	21%	78%	22%
	P	8,605,022	100%	<b>0%</b>	100%	<b>0%</b>	100%	<b>0%</b>	100%	<b>0%</b>	100%	<b>0%</b>	100%	<b>0%</b>
No pobre	A	94,883,927	78%	22%	80%	20%	28%	72%	84%	16%	75%	25%	74%	26%
	P	27,752,767	100%	<b>0%</b>	100%	<b>0%</b>	100%	<b>0%</b>	100%	<b>0%</b>	100%	<b>0%</b>	100%	<b>0%</b>

A=Ausencia  
P=Presencia

**Cuadro 4.** *Contraste de indicadores de carencia respecto la categoría de pobreza para 2016.*  
Fuente: Elaboración propia con información del CONEVAL.

Lo anterior, no es diferente para 2018 donde el 83% de la población pobre (de un total de 52,425,887 personas) presenta la carencia de seguridad social y el 76% de la población vulnerable por carencias (de un total de 36,676,375 personas) también la observa; y en el total de la población 57% presentó esta característica, es decir, en términos absolutos la población en condiciones de pobreza disminuye, pero en términos relativos se incrementa la población que tiene la carencia de derechos por seguridad social.



Categoría de Pobreza	Población	Indicador de carencia por rezago educativo		Indicador de carencia por acceso a servicios de salud		Indicador de carencia por acceso a la seguridad social		Indicador de carencia por calidad y espacios de la vivienda		Indicador de carencia de acceso a servicios básicos en la vivienda		Indicador de carencia por acceso a la alimentación		
		A	P	A	P	A	P	A	P	A	P	A	P	
Pobreza	A	72,656,529	90%	10%	87%	13%	62%	38%	96%	4%	91%	9%	89%	11%
	P	52,425,887	74%	26%	80%	20%	17%	83%	80%	20%	66%	34%	66%	34%
Vulnerable por carencias	A	88,406,041	84%	16%	88%	12%	51%	49%	88%	12%	80%	20%	80%	20%
	P	36,676,375	80%	20%	73%	27%	24%	76%	91%	9%	82%	18%	78%	22%
Vulnerable por ingresos	A	116,454,411	82%	18%	83%	17%	38%	62%	88%	12%	79%	21%	78%	22%
	P	8,628,005	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%
No pobre	A	97,730,267	78%	22%	79%	21%	27%	73%	86%	14%	75%	25%	74%	26%
	P	27,352,149	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%

A=Ausencia

P=Presencia

**Cuadro 5.** *Contraste de indicadores de carencia respecto la categoría de pobreza para 2018.*

Fuente: Elaboración propia con información del CONEVAL.

Por otro lado, para que una persona sea clasificada como pobre deben cumplirse por lo menos una carencia de derechos y no contar con ingresos superiores a la LPI, es decir que existen hogares pobres o vulnerables por carencias que pueden tener más de una carencia de derechos, incluso presentar las 6, lo cual podría profundizar la precariedad de sus condiciones de vida y sin embargo no ser categorizado como pobre porque su ingreso supera la LPI.

En este sentido, en los cuadros 6 y 7 se observa que, aunque en proporciones menores, existe población considerada como no pobre y no vulnerable por carencias pero que presenta 4, 5 y hasta 6 carencias de derechos.

	Población	Número de carencias							
		0	1	2	3	4	5	6	
Pobreza	Ausencia	69,218,543	53%	24%	17%	5%	2%	0.3%	0.0%
	Presencia	53,418,151	0%	32%	34%	21%	9%	3%	0.4%
Vulnerable por carencias	Ausencia	89,775,940	40%	19%	20%	13%	6%	2%	0.2%
	Presencia	32,860,754	0%	50%	35%	11%	3%	1%	0.1%

**Cuadro 6.** *Contraste de indicadores de carencia respecto el índice de privación, 2016.*

Fuente: Elaboración propia con información del CONEVAL.

Categoría de Pobreza		Población	Número de carencias						
			0	1	2	3	4	5	6
Pobreza	Ausencia	72,656,529	50%	25%	18%	6%	2%	0.4%	0.0%
	Presencia	52,425,887	0%	32%	34%	21%	10%	3%	0.4%
Vulnerable por carencias	Ausencia	88,406,041	41%	19%	20%	13%	6%	2%	0.2%
	Presencia	36,676,375	0%	49%	35%	11%	3%	1%	0.1%

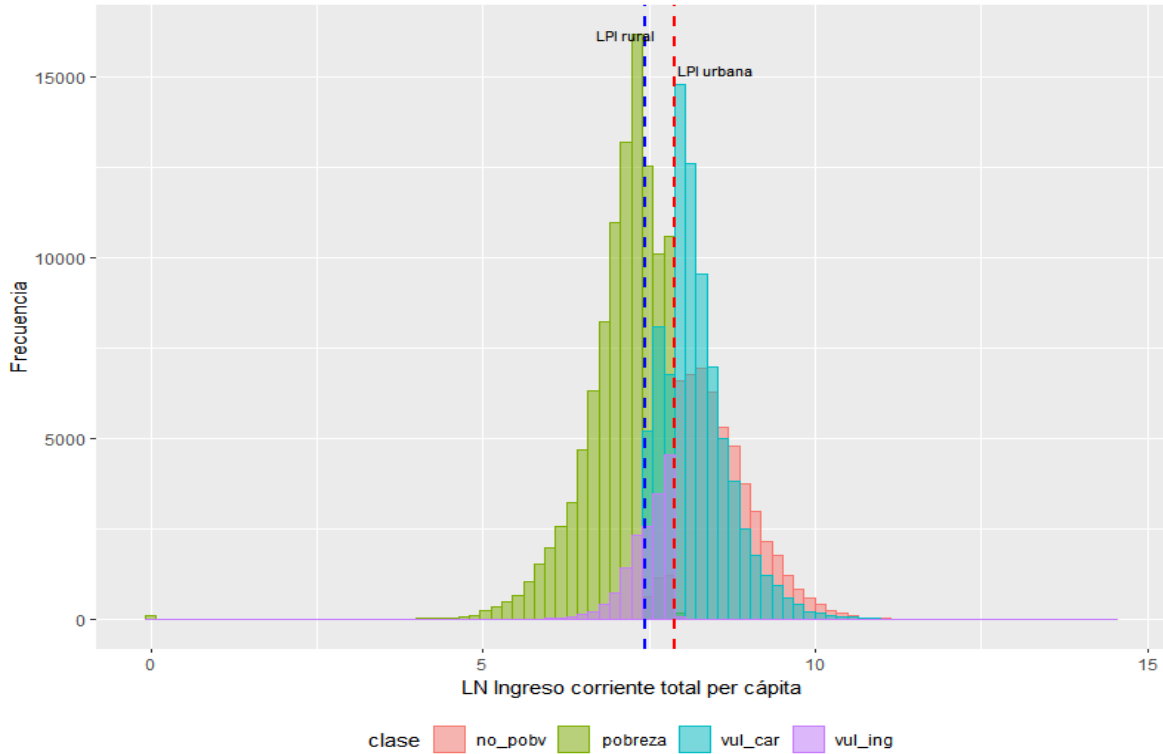
**Cuadro 7.** *Contraste de indicadores de carencia respecto el índice de privación, 2018.*

Fuente: Elaboración propia con información del CONEVAL.

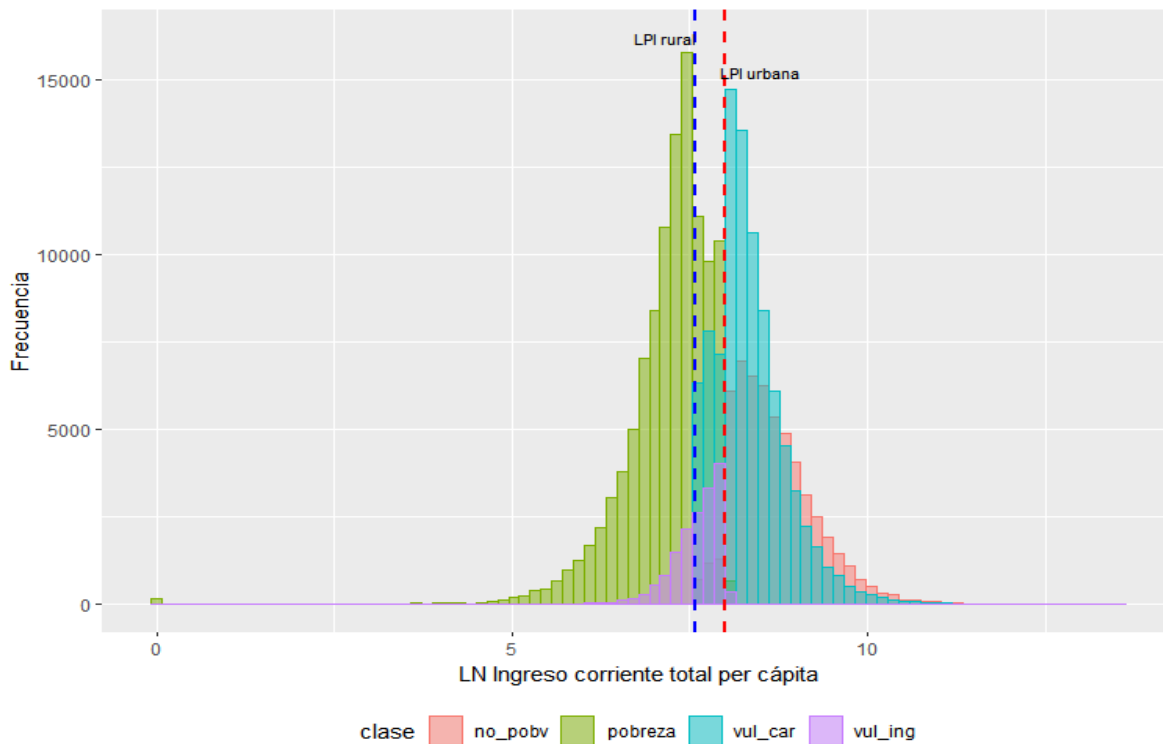
Por otro lado, el ICTPC que corresponde al estado de bienestar presenta diferentes distribuciones en cada una de las condiciones de pobreza. En los gráficos 1 y 2 se usará el ICTPC en escala logarítmica<sup>10</sup> y se añaden dos líneas verticales, una azul que representa la LPI rural y una roja que representa la LPI urbana. En los gráficos se pueden observar los siguientes fenómenos:

- La mayoría de la población se clasifica como pobre y una causa fundamental es que sus ingresos no alcanzan la LPI.
- La siguiente clasificación más frecuente es la de vulnerable por carencias, aunque se puede percibir que hay casos de personas cuyos ingresos son superiores a la LPI, razón por la cual no fueron clasificados como pobres.
- Por su parte, la población clasificada como no pobre concentra sus ingresos en niveles superiores a los ingresos de los grupos que no son clasificados como tal, con excepción de algunas personas clasificadas como vulnerables con carencias que presentan ingresos muy altos y que si presentaran la necesidad de hacer uso de algunos servicios podrían cubrirlos y satisfacerlos con sus ingresos.
- Se puede observar que para 2016 y para 2018 las distribuciones de ingresos se mantienen.

<sup>10</sup> Debido a que los datos económicos como el ingreso presentan distribuciones asimétricas, es decir la mayoría de los casos son ingresos bajos que se concentran de lado izquierdo de la curva y muy pocos se localizan en el extremo contrario al aplicar el logaritmo se minimizan dichas asimetrías (Marín Diazaraque) y esto permite una mejor apreciación gráfica.



**Gráfico 1.** *Distribución del ICTPC por condición de pobreza, 2016.*  
Fuente: Elaboración propia con información del CONEVAL.



**Gráfico 2.** *Distribución del ICTPC por condición de pobreza, 2018.*  
Fuente: Elaboración propia con información del CONEVAL.

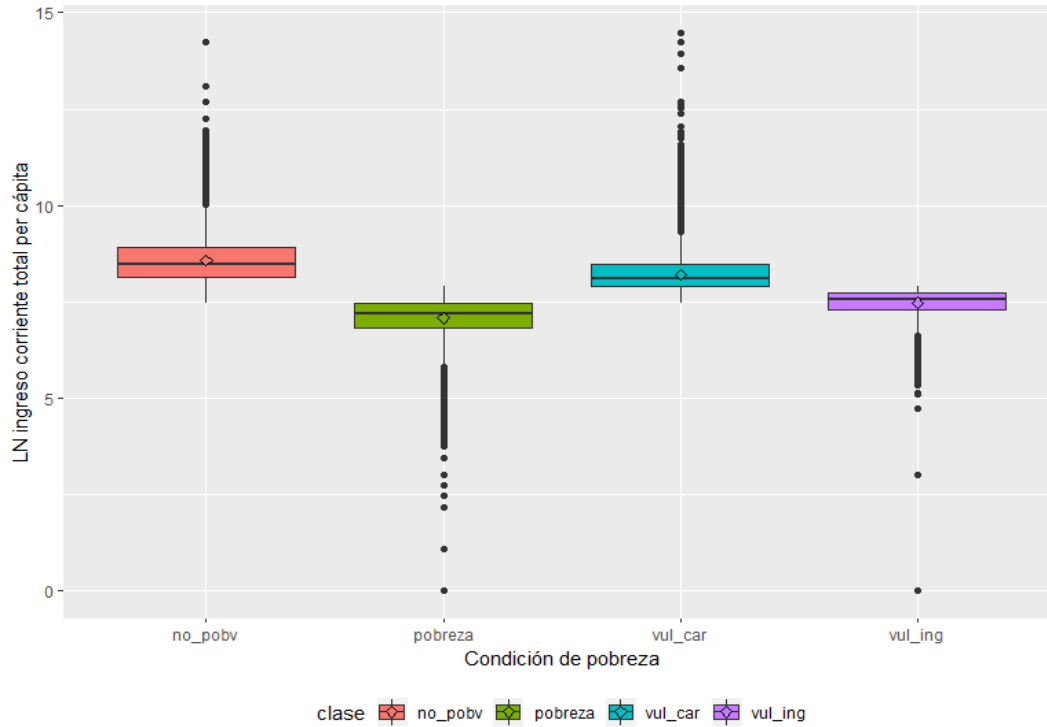
Para tener otra perspectiva de la distribución del ingreso por categoría de pobreza, a continuación, se presentan gráficos de caja, en los cuales también se usará la escala logarítmica. En los gráfico 3 y 4 se observa de forma particular en las cajas verde y lila, que corresponden respectivamente a la población clasificada como pobre y vulnerable por ingresos, que los valores atípicos se localizan en el extremo inferior; lo cual indica que existen personas cuyos ingresos son sumamente reducidos, es decir son menores a la diferencia entre el primer cuartil<sup>11</sup> y 1.5 veces el rango intercuartil<sup>12</sup>. Esto último, en contraste a lo observado en las gráficas rosa y azul, que corresponden respectivamente a los hogares clasificados como no pobres y vulnerables por carencias, donde se presentan niveles de ingresos superiores con valores atípicos en el extremo superior, lo que significa que existen personas con ingresos mayores a la suma entre el tercer cuartil<sup>13</sup> y 1.5 veces el rango intercuartil. En este sentido los hogares clasificados como no pobres tienen ingresos muy superiores a la media de los grupos que no son clasificados como tal.

---

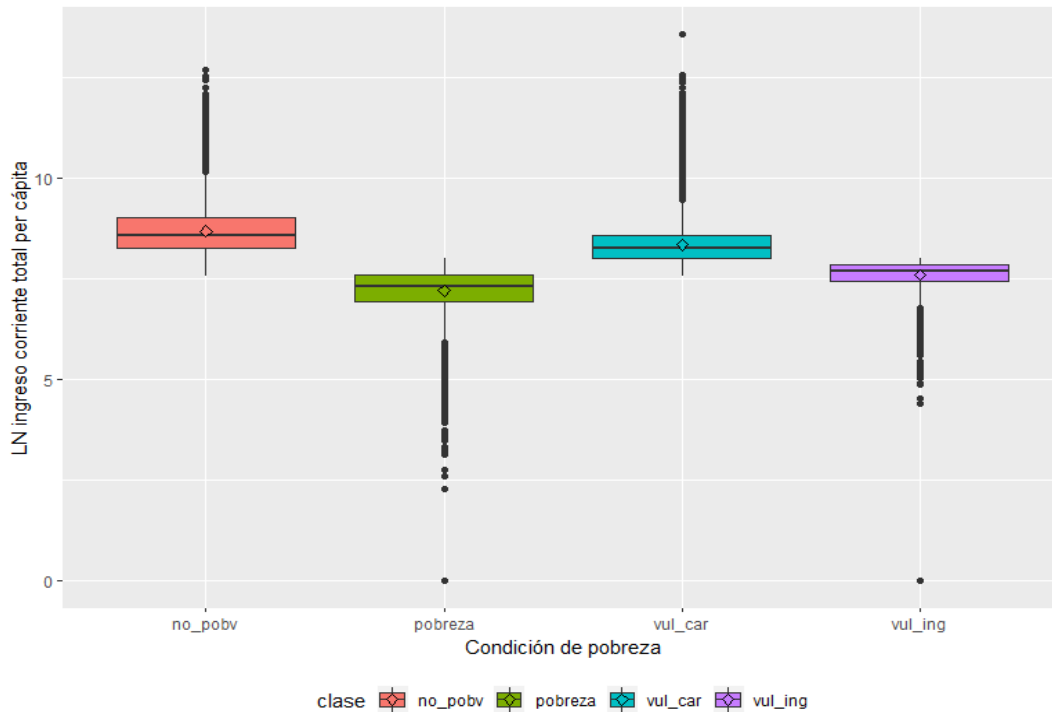
<sup>11</sup> Corresponde al valor de la observación que se encuentra en el primer cuarto de los datos ordenados de menor a mayor.

<sup>12</sup> Diferencia entre el tercer y el primer cuartil de una distribución.

<sup>13</sup> Corresponde al valor de la observación que se encuentra en el tercer cuarto de los datos ordenados de menor a mayor.

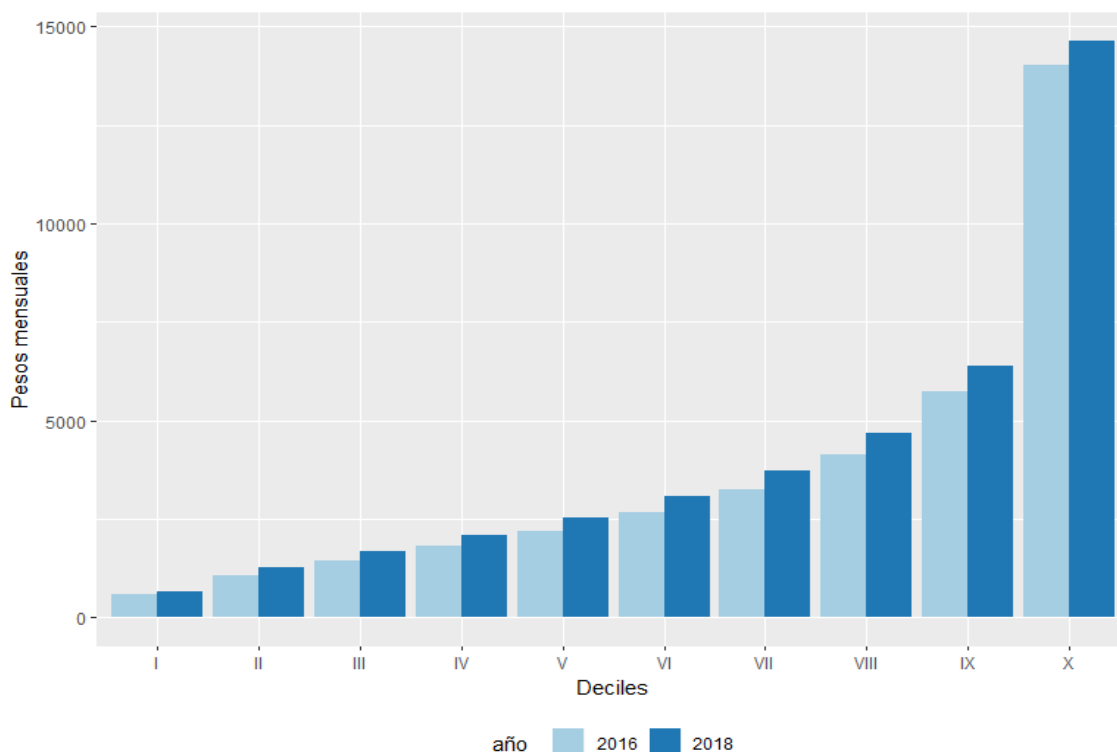


**Gráfico 3.** Gráfico de caja del ICTPC por condición de pobreza, 2016.  
Fuente: Elaboración propia con información del CONEVAL.



**Gráfico 4.** Gráfico de caja del ICTPC por condición de pobreza, 2018.  
Fuente: Elaboración propia con información del CONEVAL.

En vista de lo anterior, con la finalidad de entender mejor la variable ICTPC y poder dimensionar su magnitud, en el gráfico 5 se muestra que ordenando a la población de acuerdo a esta variable de menor a mayor y dividiéndolos en diez partes iguales, al obtener el ingreso promedio en cada una de las partes, se puede verificar que las diferencias en ingresos se acentúan entre más ingresos percibe la población. Por ejemplo, el ingreso promedio del último decil es más de dos veces superior al del noveno decil.



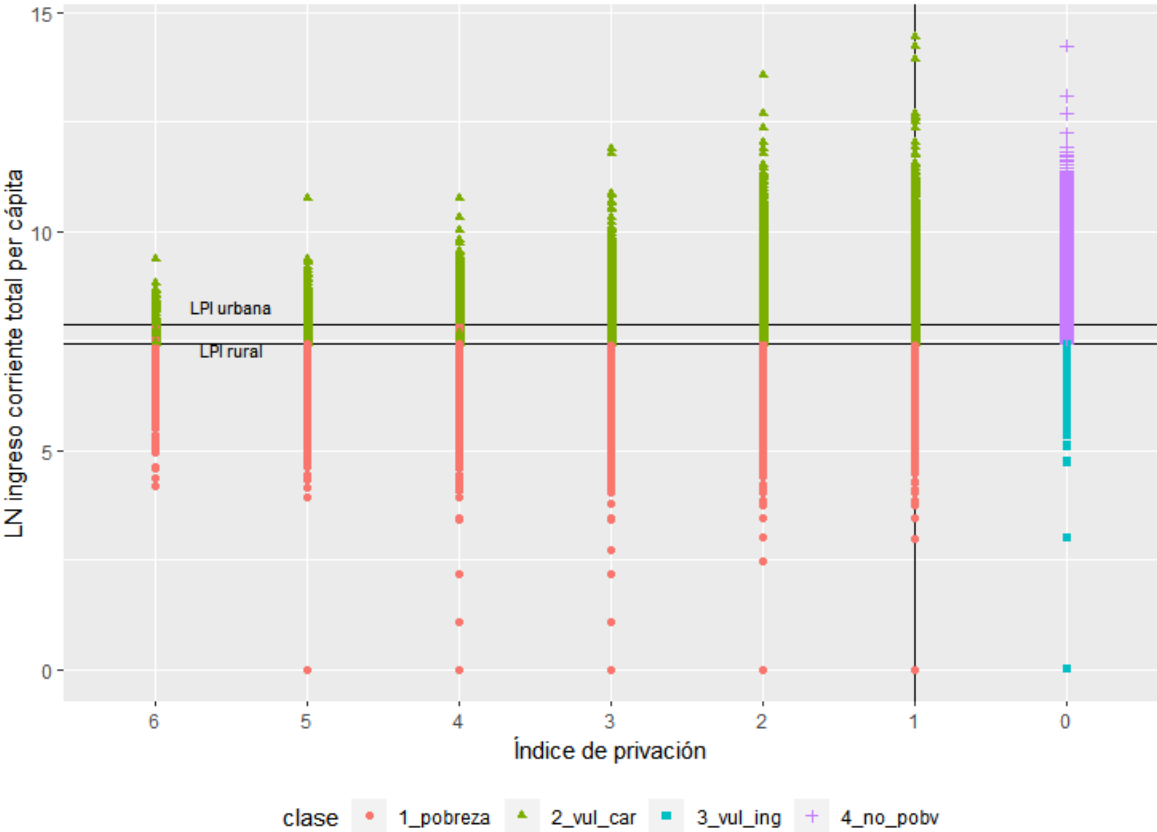
*Gráfico 5. ICTPC promedio por decil de ingresos 2016 y 2018.*  
Fuente: Elaboración propia con información del CONEVAL.

Los gráficos 6 y 7 muestran la clasificación de pobreza de acuerdo con la metodología del CONEVAL. Es decir, se considera la partición en los espacios de bienestar (delimitados por la LPI<sup>14</sup>) medido en el eje vertical, y el espacio de derechos (Índice de privación) el cual se mide en el eje horizontal, dónde se pueden identificar el número de carencias que enfrenta cada persona. De esta manera, se puede observar que existen personas en condición de pobreza (con

<sup>14</sup> Se contemplan una LPI para el ámbito urbano y otra para el ámbito rural.

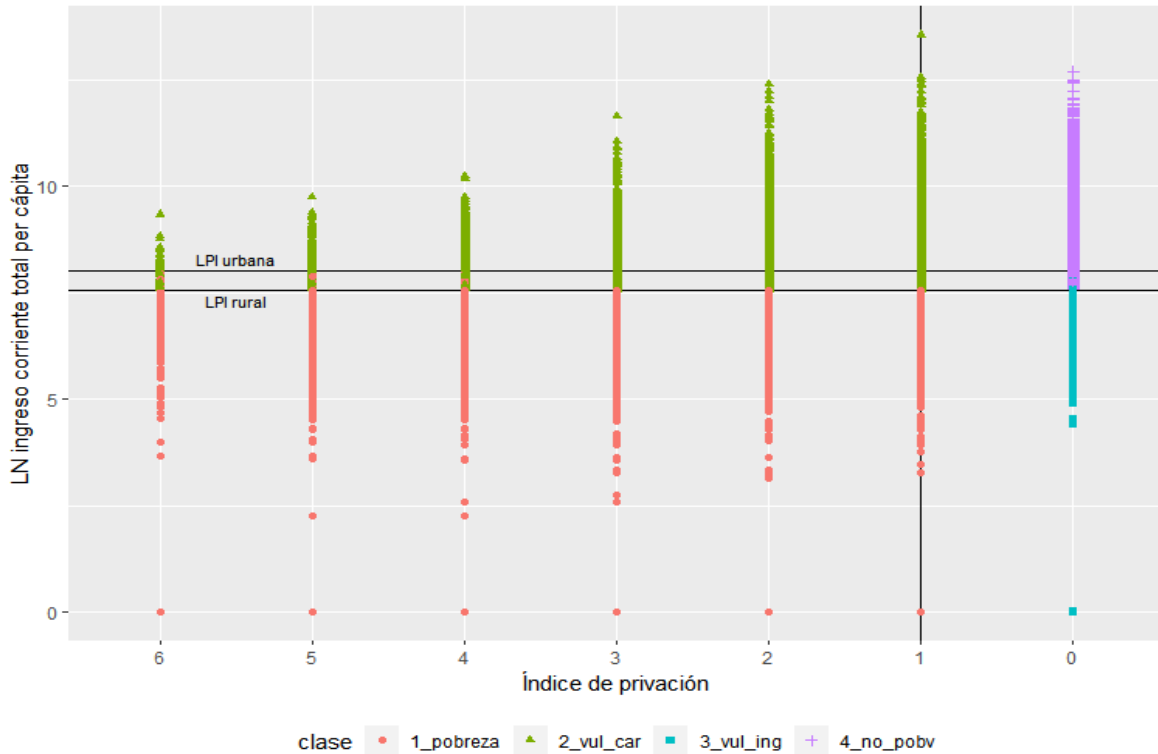
ingresos por debajo de la LPI) que además se enfrentan a un número importante de carencias. Por definición, el CONEVAL considera pobreza extrema cuando existen más de 3 carencias y el ingreso se encuentra por debajo de la línea de pobreza extrema por ingresos<sup>15</sup>.

En los gráficos las dos líneas horizontales representan la LPI urbana y rural, que el CONEVAL identifica por separado, estas líneas marcan el ingreso corriente total que una persona necesita para cubrir sus necesidades básicas. La línea vertical representa el índice de privación cuando es igual a uno, lo que significa que los puntos a la izquierda de esta línea, representan a personas que tienen por lo menos una carencia de derechos.



**Gráfico 6.** Clasificación de pobreza de acuerdo con la metodología del CONEVAL, 2016. Fuente: Elaboración propia con información del CONEVAL.

<sup>15</sup> Valor monetario de una canasta alimentaria básica (CONEVAL, s.f.).



**Gráfico 7.** Clasificación de pobreza de acuerdo con la metodología del CONEVAL, 2018.  
Fuente: Elaboración propia con información del CONEVAL.

### 3.2 Parte 2. Una clasificación alternativa de la pobreza

En este apartado se presentan los resultados obtenidos después de someter los mismos datos que usó el CONEVAL para hacer sus clasificaciones de pobreza, a algoritmos de clasificación (aprendizaje supervisado) y agrupación (aprendizaje no supervisado) y de esta forma identificar clases o grupos de pobreza o no pobreza alternativos.

En primer lugar, se optó por usar técnicas de aprendizaje supervisado, ya que de acuerdo con los datos disponibles se contaba con los insumos necesarios para entrenar estos algoritmos ya que se tenían tanto los atributos como las variables objetivo. En este sentido, se usaron las técnicas de regresión logística multinomial y Naïve Bayes debido a que tienden a generar buenos resultados de clasificación y sus niveles de complejidad son reducidos, es decir suelen ofrecer simplicidad y efectividad.



En referencia a la técnica de agrupamiento de aprendizaje no supervisado, fue necesario localizar una alternativa que pudiera combinar el uso de datos categóricos con datos numéricos, ya que la técnica más común que es k-means se alimenta de datos numéricos. Por lo cual, se usó K-prototypes que está diseñada precisamente para combinar datos numéricos y categóricos.

### **3.2.1 Regresión logística multinomial**

La regresión logística forma parte del grupo de algoritmos de clasificación en aprendizaje supervisado. En éstos se entrena el algoritmo para que aprenda de la interacción entre los atributos y una variable objetivo y de este aprendizaje pueda determinar las clases a las que pertenece cada registro. Este algoritmo tiene como objetivo modelar la probabilidad posterior, es decir la probabilidad de que un registro se clasifique en determinada clase dado que existe evidencia que otro evento sucede (por ejemplo, en este caso la clasificación de pobreza dado que exista la carencia de seguridad social). El algoritmo toma como entrada el número K de clases y para cada registro asigna un valor de probabilidad entre 0 y 1, que indicará la posibilidad de que se clasifique en una clase o en otra (Hastie, Tibshirani, & Friedman, 2008).

Cuando  $K=2$ , se dice que se tiene una variable objetivo binaria. En este caso el modelo es simple al tratarse de una función lineal y puede ser muy útil para clasificaciones médicas como si el paciente tiene cierta “enfermedad” o no. Es decir, para decidir si cierta característica está presente o ausente (Hastie, Tibshirani, & Friedman, 2008).

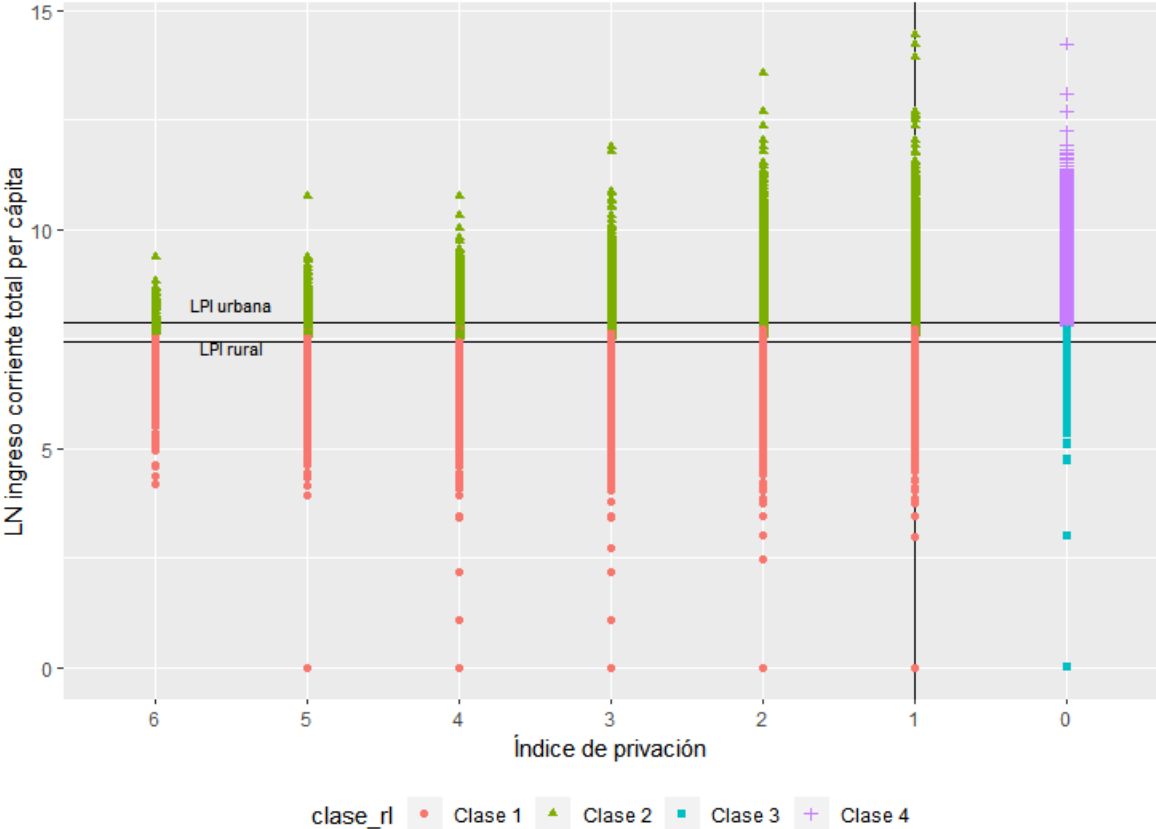
En el caso del modelo general, para  $K>2$ , la regresión logística se usa para clasificar variables que pueden seguir una distribución multinomial. En particular, se crea un modelo de regresión lineal para cada clase y luego se combina en una sola ecuación asegurando que la restricción de probabilidad de la suma de todas las probabilidades sea 1 (Ramasubramanian & Singh, 2019).

Este algoritmo acepta variables categóricas y numéricas y a partir de dicha información se obtiene la probabilidad de cada registro por cada clase lo que

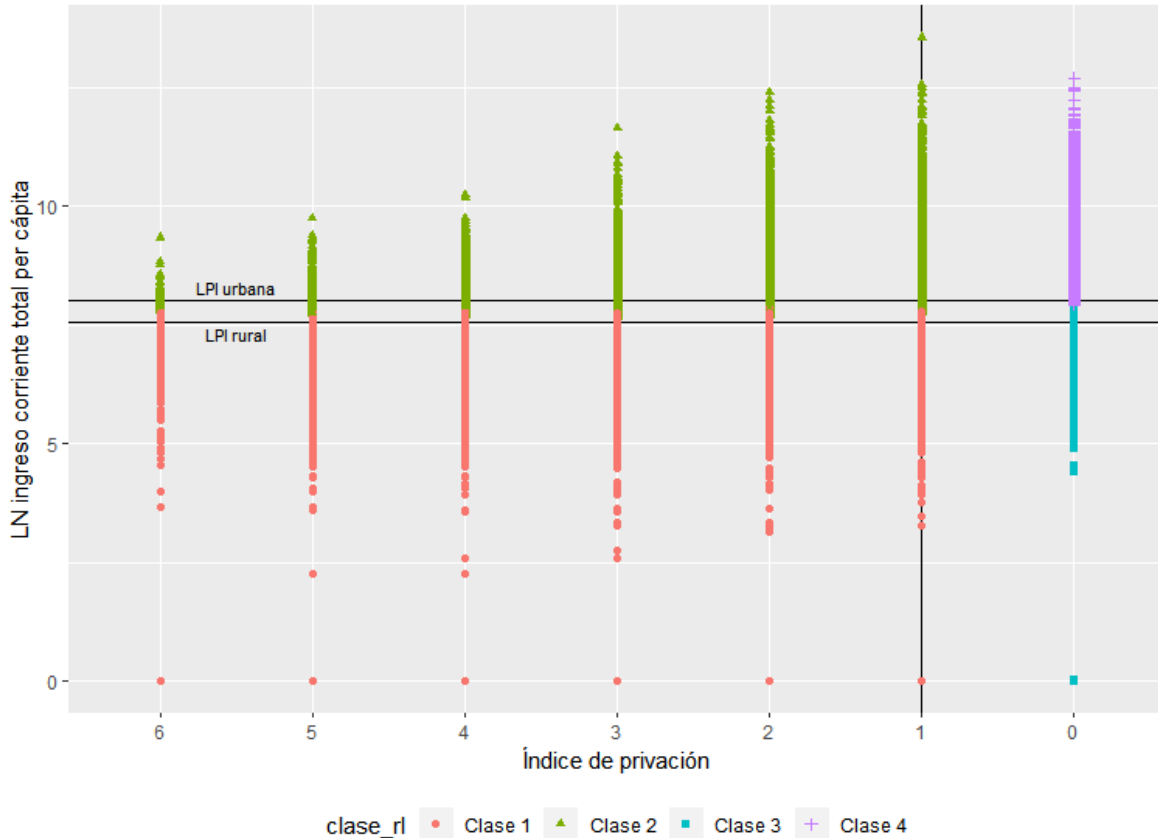
permite hacer una predicción de para decidir si dicho registro se clasifica en una u otra.

**3.2.1.1 Resultados**

A continuación, en los gráficos 8 y 9 se presentan los resultados obtenidos de la clasificación hecha mediante regresión logística multinomial para los años 2016 y 2018. Estos gráficos se presentan de la misma forma que aquellos en los que se presentó la clasificación del CONEVAL (gráficos 6 y 7), con la finalidad de facilitar la comparación de los resultados, es decir el eje horizontal mide el índice de privación y en el eje vertical se mide el ICTPC en términos logarítmicos, para su mejor apreciación.



**Gráfico 8.** Clasificación de la condición de pobreza predicha por la regresión logística multinomial, 2016.  
Fuente: Elaboración propia con información del CONEVAL.



**Gráfico 9.** Clasificación de la condición de pobreza predicha por la regresión logística multinomial, 2018.

Fuente: Elaboración propia con información del CONEVAL.

La matriz de confusión actúa como una tabla de contingencia que contribuye a identificar los registros que fueron clasificados o predichos correctamente comparados con cada clase de los datos originales. Por lo cual contribuye como herramienta para identificar las coincidencias o diferencias de clasificación entre la metodología del CONEVAL y la realizada por el algoritmo.

Categoría de Pobreza/ Clase	Clase 1	Clase 2	Clase 3	Clase 4
<b>Pobreza</b>	<b>96,783</b>	<b>8,983</b>	<b>0</b>	<b>0</b>
<b>Vulnerable por carencias</b>	<b>11,585</b>	<b>69,764</b>	<b>0</b>	<b>0</b>
<b>Vulnerable por ingresos</b>	<b>0</b>	<b>0</b>	<b>13,823</b>	<b>2,438</b>
<b>No pobre ni vulnerable</b>	<b>0</b>	<b>0</b>	<b>2,387</b>	<b>51,884</b>

**Cuadro 8.** Matriz de confusión resultados CONEVAL respecto regresión logística multinomial, 2016.

Fuente: Elaboración propia con información del CONEVAL.

<b>Categoría de Pobreza/ Clase</b>	<b>Clase 1</b>	<b>Clase 2</b>	<b>Clase 3</b>	<b>Clase 4</b>
<b>Pobreza</b>	<b>97,874</b>	<b>9,684</b>	<b>0</b>	<b>0</b>
<b>Vulnerable por carencias</b>	<b>12,360</b>	<b>77,589</b>	<b>0</b>	<b>0</b>
<b>Vulnerable por ingresos</b>	<b>0</b>	<b>0</b>	<b>13,407</b>	<b>2,567</b>
<b>No pobre ni vulnerable</b>	<b>0</b>	<b>0</b>	<b>2,392</b>	<b>53,183</b>

**Cuadro 9.** *Matriz de confusión resultados CONEVAL respecto regresión logística multinomial, 2018.*

Fuente: Elaboración propia con información del CONEVAL.

### 3.2.1.2 Conclusiones

Se observa que usando las mismas variables que emplea el CONEVAL para la clasificación de pobreza multidimensional y aplicando un algoritmo de regresión logística multinomial que clasifica la población en cuatro clases distintas, al comparar ambos resultados para 2016 y 2018 se obtiene una exactitud de 0.9014 y 0.8996, respectivamente. Las diferencias entre los resultados del algoritmo y la clasificación del CONEVAL se distribuyen a lo largo de todas las clases.

Cabe señalar que asumiendo que la clase 1 corresponde a los pobres, la clase 2 a los vulnerables por carencias, la clase 3 a los vulnerables por ingresos y la clase 4 a los no pobres; las principales diferencias entre los resultados de la clasificación del CONEVAL y la clasificación de la regresión logística multinomial se dan entre las clase 1 y 2, es decir, el algoritmo clasifica como pobres a personas que la metodología diseñada por el CONEVAL calificaría como personas con vulnerabilidad de carencias y viceversa. Del mismo modo se observa que existen inversiones de clasificación entre las clases 3 y 4 que corresponde a los a personas con vulnerabilidad de ingresos y no pobres.

En este sentido, el algoritmo no logra identificar la doble condicionante, que implica ingresos menores a la LPI y presentar por lo menos una carencia de derechos, para la categorización de pobreza que implementa el CONEVAL e identifica que, de acuerdo a los atributos utilizados en el algoritmo, algunos de los hogares en condición de pobreza y vulnerables por carencias tienden a ser más parecidos entre ellos; el mismo fenómeno ocurre entre los hogares vulnerables por

ingresos y los no pobres. De acuerdo con dichos resultados se puede inferir que la doble condición resulta estricta para personas que, pese a tener condiciones de vida complejas al presentar distintas carencias en cuanto a derechos básicos no son contemplados como pobres, debido a que apenas sus ingresos superan la LPI.

### 3.2.2 Naïve Bayes

Este algoritmo también pertenece al grupo de aprendizaje supervisado y es apropiado cuando la dimensionalidad de las variables es alta. El principal supuesto es que las variables que servirán de entrada para entrenar al algoritmo son independientes entre ellas, sin embargo puede ser que dicha consideración no se satisfaga completamente y aun así obtener buenos resultados (Bishop, 2006) .

La probabilidad condicional juega un papel importante en determinar el impacto de un evento en otro. Intuitivamente, una probabilidad condicional es la consideración de que la probabilidad de un evento podría aumentar o disminuir si se sabe que otro evento influye en el evento en estudio (Ramasubramanian & Singh, 2019). El teorema de Bayes define la probabilidad condicional entre dos eventos:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Donde:

$P(A)$  es la probabilidad a priori de A.

$P(A|B)$  es la probabilidad a posteriori y puede ser leída como la probabilidad de A dado B.

$P(B)$  es la probabilidad total del evento B.

$P(B|A)$  es la probabilidad de B dado el evento A.

$P(B|A)P(A)$  puede describirse como probabilidad conjunta donde los eventos A y B se intersectan, es decir, es la probabilidad de que A y B sucedan juntos.

La expresión puede ser simplificada, ya que para un mismo ejemplo todos los eventos tienen la misma probabilidad marginal  $P(B)$ . Por lo tanto, se puede omitir, obteniendo una expresión proporcional:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \propto P(B|A)P(A)$$

En nuestro caso particular, con los datos del CONEVAL, este modelo se puede interpretar como la probabilidad de categorización de pobreza dado determinado indicador de derechos o de bienestar. Así, para la determinación de la clase de cada individuo se determinará la clase que obtenga la probabilidad máxima, considerando la probabilidad de cada característica en la clase y ponderado por la probabilidad de la clase.

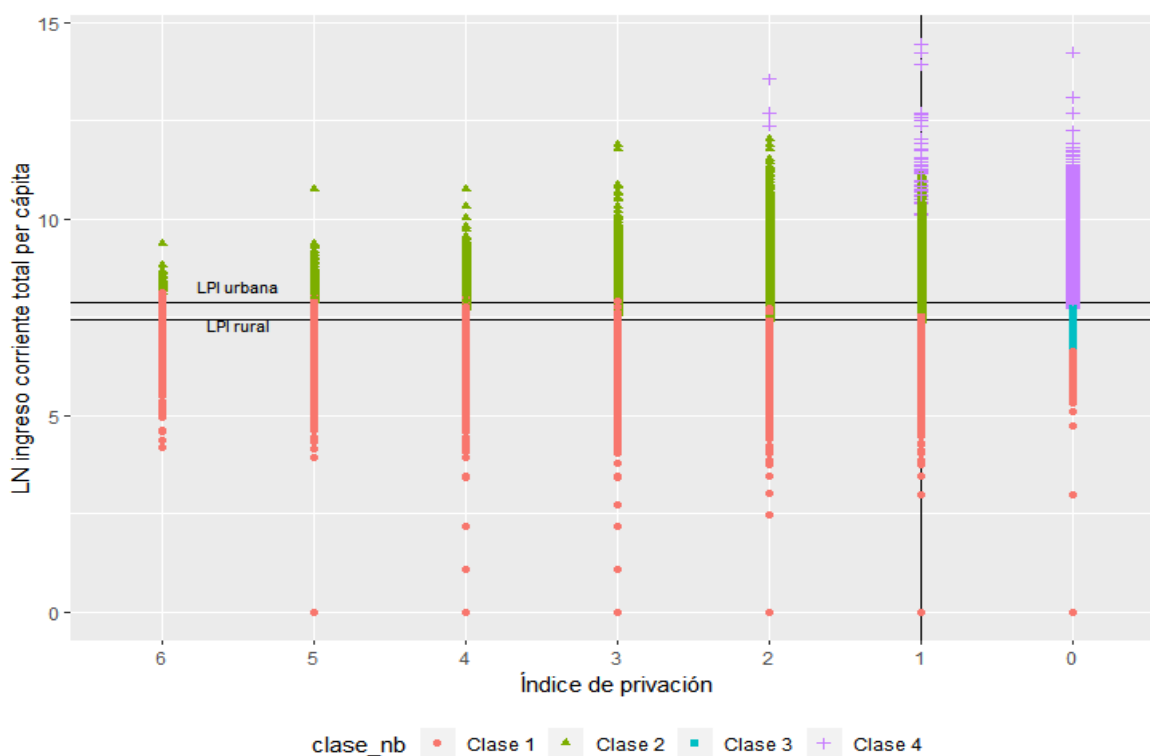
$$P(C_k|x) = P(x|C_k)P(C_k)$$

El algoritmo puede ser implementado con variables categóricas y variables numéricas debido a que cada una puede ser representada separadamente usando los modelos apropiados, por ejemplo usar distribuciones Bernoulli para observaciones binarias y Gaussianas para datos con distribuciones continuas (Bishop, 2006).

### 3.2.2.1 Resultados

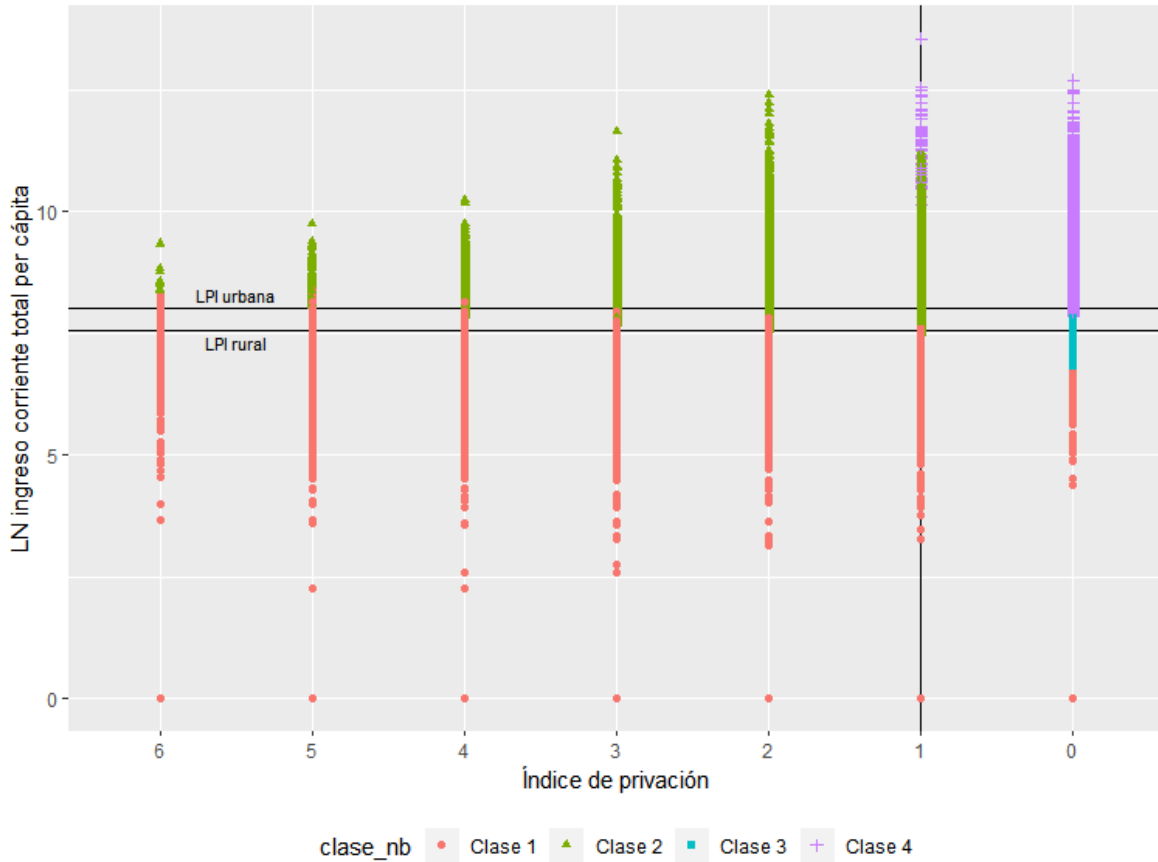
Los resultados obtenidos de la clasificación hecha mediante Naïve Bayes se presentan mediante los gráficos 10 y 11, para los años 2016 y 2018, respectivamente. Una vez más, la clasificación hecha por el algoritmo Naïve Bayes se presenta en el mismo formato que aquella del CONEVAL (gráficos 6 y 7). En este caso aplicamos el logaritmo natural a la variable ICTPC antes de aplicar el algoritmo. De esta forma se obtienen mejores resultados, lo anterior es

consecuencia de que al aplicar el logaritmo la distribución de dicha variable se apegue más a la de una distribución normal.<sup>16</sup>



**Gráfico 10.** Clasificación de la condición de pobreza predicha por Naïve Bayes, 2016.  
Fuente: Elaboración propia con información del CONEVAL.

<sup>16</sup> Es común que los datos económicos como el ingreso presenten distribuciones asimétricas, ya que la mayoría de las personas observan bajos niveles de ingresos, lo cual en términos de distribución concentran los datos de lado izquierdo de la curva; al aplicar el logaritmo se minimizan dichas asimetrías y se obtiene una mejor distribución de los datos, lo cual contribuye a mejorar sus resultados (Marín Diazaraque). Por otro lado es común localizar el logaritmo como herramienta para medición de brechas de ingresos como es el Índice Watts (UNECE, 2017).



**Gráfico 11.** Clasificación de la condición de pobreza predicha por Naïve Bayes, 2018.  
Fuente: Elaboración propia con información del CONEVAL.

Además de los gráficos, a continuación, se presentan una comparación tabular de los resultados por clase contra la clasificación del CONEVAL.

Categoría de Pobreza/ Clase	Clase 1	Clase 2	Clase 3	Clase 4
<b>Pobreza</b>	92,226	13,540	0	0
<b>Vulnerable por carencias</b>	14,125	67,155	0	69
<b>Vulnerable por ingresos</b>	493	0	11,512	4,256
<b>No pobre ni vulnerable</b>	0	0	1,867	52,404

**Cuadro 10.** Matriz de confusión resultados CONEVAL respecto Naïve Bayes, 2016.  
Fuente: Elaboración propia con información del CONEVAL.



<b>Categoría de Pobreza/ Clase</b>	<b>Clase 1</b>	<b>Clase 2</b>	<b>Clase 3</b>	<b>Clase 4</b>
<b>Pobreza</b>	92,009	15,549	0	0
<b>Vulnerable por carencias</b>	13,981	75,878	0	90
<b>Vulnerable por ingresos</b>	471	0	11,088	4,415
<b>No pobre ni vulnerable</b>	0	0	1,833	53,742

**Cuadro 11.** *Matriz de confusión resultados CONEVAL respecto Naïve Bayes, 2018.*  
Fuente: Elaboración propia con información del CONEVAL.

### 3.2.2.2 Conclusiones

Utilizando las mismas variables que usa el CONEVAL para la clasificación de pobreza multidimensional y aplicando el logaritmo natural al ICTPC para realizar la clasificación en cuatro clases, se obtienen exactitudes de 0.8667 y 0.8649 para 2016 y 2018, respectivamente.

Cabe señalar que como se asumió anteriormente, la clase 1 es comparable a quienes son clasificados en condición de pobreza, la clase 2 a los vulnerables por carencias, la clase 3 a los vulnerables por ingresos y la clase 4 a los no pobres; las principales diferencias entre los resultados de la clasificación del CONEVAL y la clasificación de Naïve Bayes se presentan, al igual que en los resultados de clasificación de la regresión logística multinomial, entre las clase 1 y 2, es decir, el algoritmo clasifica como pobres a personas que la metodología diseñada por el CONEVAL definiría como personas con vulnerabilidad de carencias y viceversa.

También existe una clasificación invertida entre las clases 3 y 4 que corresponde a los a personas con vulnerabilidad de ingresos y no pobres ni vulnerables. En este sentido el algoritmo también agrupa en una misma clase a las personas con montos de ingresos altos y con ninguna o muy pocas carencias de derechos, lo cual permite deducir que este algoritmo podría detectar que los ingresos altos pueden ayudar cuando se tienen carencias, y no precisamente se es vulnerable.

En conclusión, se observa que el algoritmo no logra identificar la doble condicionante, que implica no contar con ingresos superiores a la LPI y presentar por lo menos una carencia de derechos, para la categorización de pobreza que implementa el CONEVAL e identifica que, de acuerdo a los atributos utilizados, algunos de los hogares en condición de pobreza y vulnerables por carencias tienden a ser más parecidos entre ellos. El mismo fenómeno ocurre entre los hogares vulnerables por ingresos y los no pobres ni vulnerables.

### **3.2.3 Agrupamiento (Clustering)**

Este tipo de algoritmos corresponde al conjunto de algoritmos de aprendizaje no supervisado, donde no se cuenta con una variable objetivo. A diferencia del aprendizaje supervisado, que cuenta con un “oráculo” que es el conjunto de entrenamiento del cual aprende, el aprendizaje no supervisado no tiene dicho “oráculo”. En este sentido, en el aprendizaje supervisado es más sencillo identificar si las predicciones fueron exitosas o no, sin embargo, cuando se trata de aprendizaje no supervisado las validaciones de efectividad del algoritmo son más complejas (Hastie, Tibshirani, & Friedman, 2008).

Entre los algoritmos de aprendizaje no supervisado se ubican los algoritmos de agrupamiento (clustering) cuyo objetivo es segmentar los registros en grupos, de modo que las diferencias entre los asignados al mismo grupo tienden a ser más pequeñas que entre aquellos que se asignaron a grupos diferentes (Hastie, Tibshirani, & Friedman, 2008).

Existen diferentes algoritmos de partición, entre los cuales “K-means” es uno de los más conocidos. Éste se basa en seleccionar aleatoriamente los objetos<sup>17</sup> iniciales para identificar los más cercanos a éstos y así determina los primeros grupos, después determina la media de los grupos previamente definidos y recalcula la similitud entre cada objeto y dicha media y reubica a cada objeto en el grupo cuya media es más cercana, este proceso se realiza iterativamente. La

---

<sup>17</sup> En este caso un objeto corresponderá a los atributos del espacio de bienestar y de derechos de una persona.

meta del proceso es hacer que la sumatoria de todas las distancias respecto de su media sea la mínima posible (Huang, 1997).

K-means funciona de forma muy eficiente cuando se trata de datos numéricos. Sin embargo, no necesariamente funciona bien con datos categóricos o con la combinación de ambos. En esos casos se puede hacer uso de un algoritmo alternativo cuyo fundamento se encuentra en K-means pero con adecuaciones. Cuando se trata de analizar una mezcla de datos numéricos y categóricos, se puede utilizar el algoritmo K-prototypes. El cual integra K-means para datos numéricos y K-modes para datos categóricos.

K-modes es un algoritmo que únicamente considera datos categóricos y presenta los siguientes cambios respecto de K-means: a) usa una medida de discrepancias entre datos categóricos; b) reemplaza el uso de la media por la moda; y c) usa la frecuencia para localizar la modas en cada grupo, la medida de similitud se determina por el número de discrepancias entre dos objetos categóricos, así el menor número de discrepancias significa mayor similitud entre los objetos (Huang, 1998).

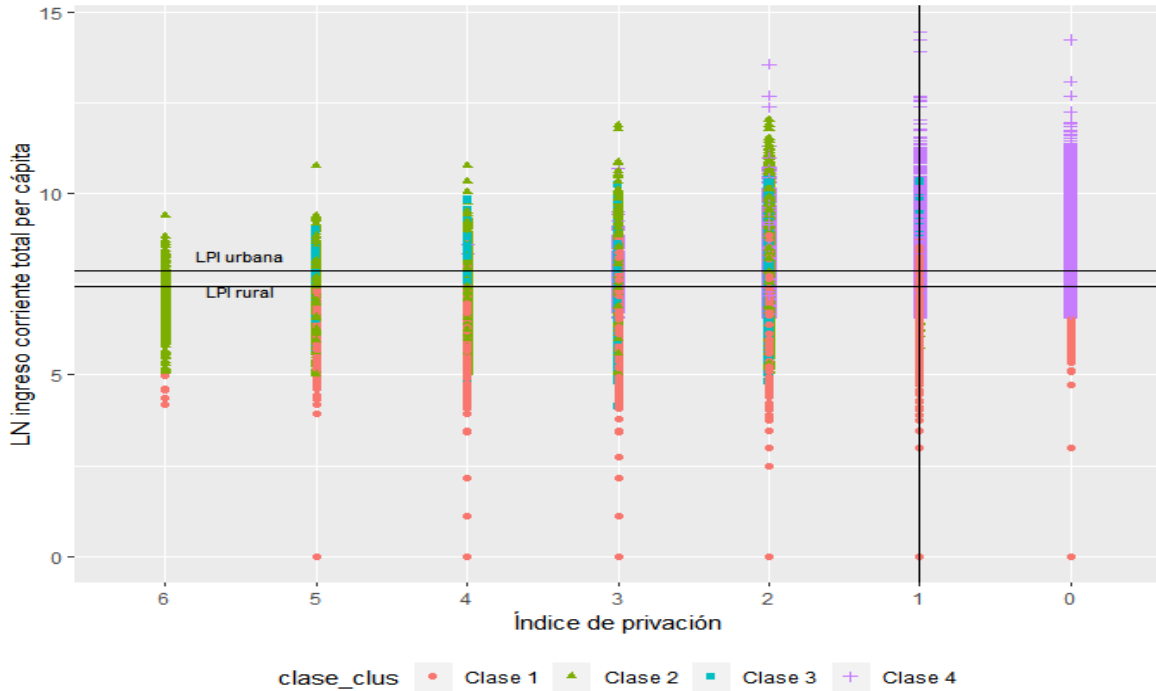
En el algoritmo K-prototypes la medida de similitud para la parte que corresponde a los datos numéricos se define simplemente por la distancia euclidiana al cuadrado, y sigue la misma lógica iterativa de K-means (Huang, 1997).

### **3.2.3.1 Resultados**

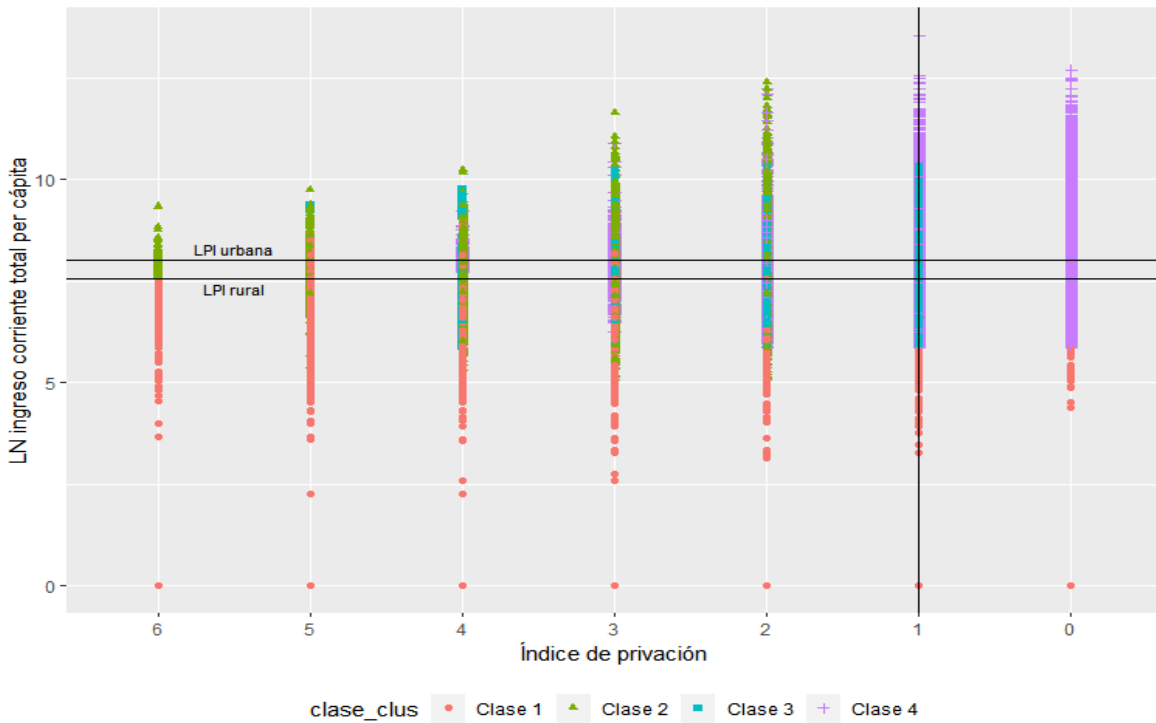
En los gráficos 12 y 13 se presentan los resultados de aplicar el algoritmo K-prototypes al mismo conjunto de datos utilizados por el CONEVAL, pero se transforma el ICTPC en términos logarítmicos.<sup>18</sup> Con esta clasificación, los resultados divergen considerablemente respecto de la clasificación del CONEVAL.

---

<sup>18</sup> Esta transformación también se implementó en Naïve Bayes.



**Gráfico 12.** Clasificación de la condición de pobreza predicha con algoritmos de agrupamiento, 2016  
 Fuente: Elaboración propia con información del CONEVAL.



**Gráfico 13.** Clasificación de la condición de pobreza predicha con algoritmos de agrupamiento, 2018.  
 Fuente: Elaboración propia con información del CONEVAL.

A continuación, se presentan los resultados del contraste de categorías de pobreza resultantes de la metodología del CONEVAL y la agrupación hecha con K-prototypes.

<b>Categoría de Pobreza/ Clase</b>	<b>Clase 1</b>	<b>Clase 2</b>	<b>Clase 3</b>	<b>Clase 4</b>
<b>Pobreza</b>	68,520	17,842	7,480	11,924
<b>Vulnerable por carencias</b>	29,622	17,231	15,956	18,540
<b>Vulnerable por ingresos</b>	356	0	0	15,905
<b>No pobre ni vulnerable</b>	0	0	0	54,271

**Cuadro 12.** *Matriz de confusión resultados CONEVAL respecto algoritmos de agrupamiento, 2016.*

Fuente: Elaboración propia con información del CONEVAL.

<b>Categoría de Pobreza/ Clase</b>	<b>Clase 1</b>	<b>Clase 2</b>	<b>Clase 3</b>	<b>Clase 4</b>
<b>Pobreza</b>	40,627	13,372	40,658	12,901
<b>Vulnerable por carencias</b>	12,251	18,865	36,451	22,382
<b>Vulnerable por ingresos</b>	44	0	0	15,930
<b>No pobre ni vulnerable</b>	0	0	0	55,575

**Cuadro 13.** *Matriz de confusión resultados CONEVAL respecto algoritmos de agrupamiento, 2018.*

Fuente: Elaboración propia con información del CONEVAL.

### 3.2.3.2 Conclusiones

De acuerdo con el cuadro 12, en 2016 la **clase 1** agrupa a personas que originalmente el CONEVAL ha clasificado como pobre, así como personas vulnerables por carencias, por ingresos y no agrupa a no pobres ni vulnerables. Por otro lado, se observa en el cuadro 14 que esta clase agrupa principalmente a las personas que presentan una combinación importante de carencias de acceso a la seguridad social y de servicios básicos en la vivienda.

En el cuadro 12 se puede observar que la **clase 2** agrupa a personas pobres, así como personas vulnerables por carencias, y en el cuadro 14 se

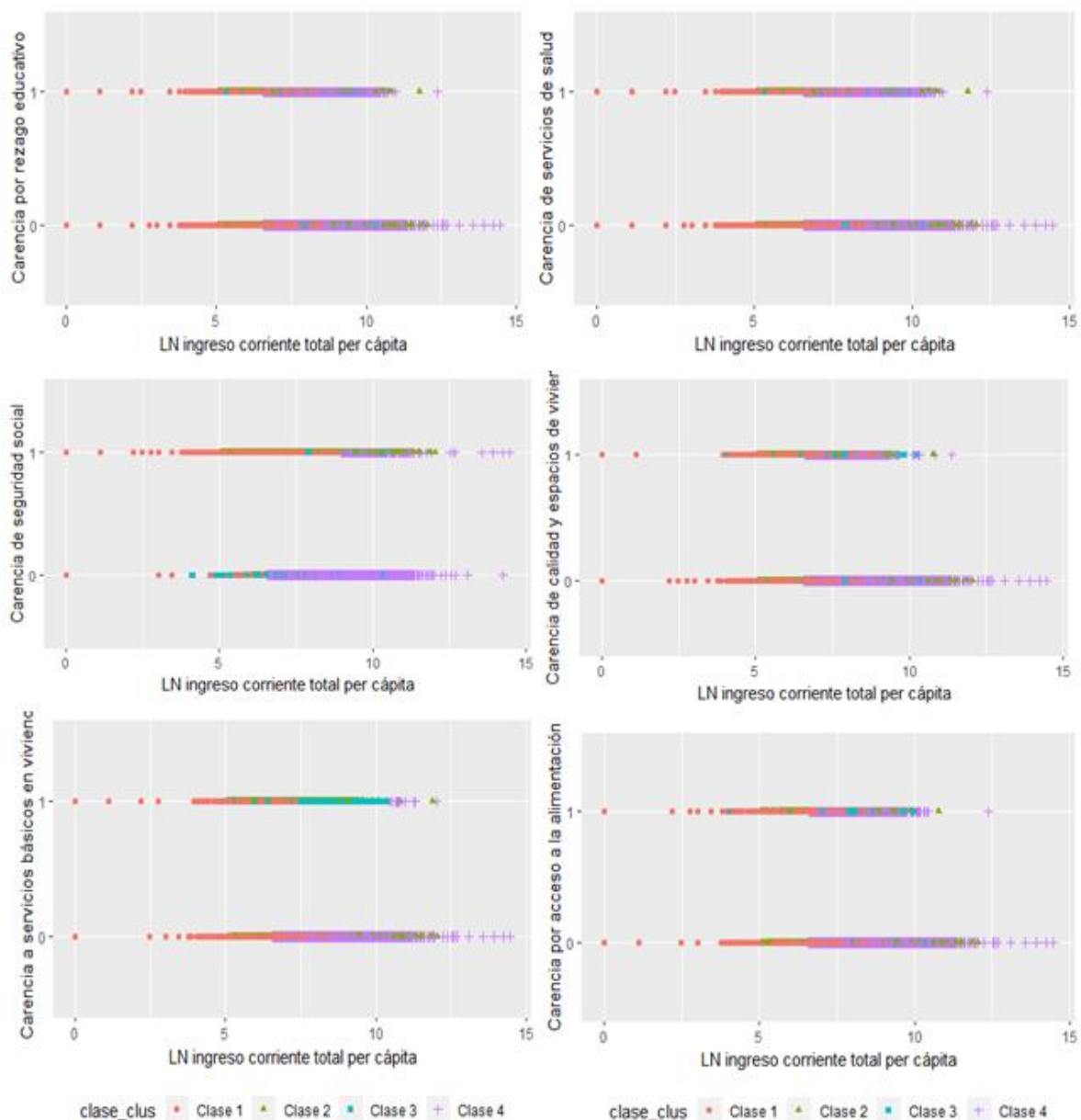
percibe como estas personas tienen la carencia de acceso a servicios a la salud y acceso a seguridad social.

La **clase 3**, de acuerdo con el cuadro 12, concentra a personas que originalmente son considerados pobres y vulnerables por carencias y que, de acuerdo con el cuadro 14, presentan la característica de no tener acceso a los servicios básicos de vivienda.

Clase	Indicador de carencia por rezago educativo	Indicador de carencia por acceso a los servicios de salud	Indicador de carencia por acceso a la seguridad social	Indicador de carencia por calidad y espacios de la vivienda	Indicador de carencia por acceso a los servicios básicos en la vivienda	Indicador de carencia por acceso a la alimentación
<b>Clase 1</b>	20%	0%	99%	18%	28%	26%
<b>Clase 2</b>	21%	100%	98%	13%	22%	23%
<b>Clase 3</b>	31%	1%	44%	21%	100%	26%
<b>Clase 4</b>	13%	1%	3%	5%	0%	14%

**Cuadro 14.** *Proporción de carencias presentadas respecto del total de población en cada clúster, 2016.*

Fuente: Elaboración propia con información del CONEVAL.



**Gráfico 14.** Diferencias identificadas en los grupos por el contraste de carencias de derechos y el ingreso, 2016.

Fuente: Elaboración propia con información del CONEVAL.

De acuerdo con el cuadro 13, en 2018 la **clase 1** contempla a personas que el CONEVAL clasificó como pobres, así como a personas vulnerables por carencias y por ingresos, en este último caso con menor presencia. Por otro lado, se observa en el cuadro 15 que esta clase agrupa principalmente a las personas que presentan una combinación de carencias de acceso a la seguridad social y de servicios básicos en la vivienda.

En el cuadro 13 se puede ver que la **clase 2** agrupa a personas pobres y vulnerables por carencias y en el cuadro 15 se percibe que una importante proporción de estas personas tienen la carencia de acceso a servicios a la salud y acceso a seguridad social.

En la **clase 3** se concentra a personas que originalmente son considerados como pobres y vulnerables por carencias, ver el cuadro 13, además de que presentan la carencia de acceso a la seguridad social en el 100% de los casos, ver el cuadro 15.

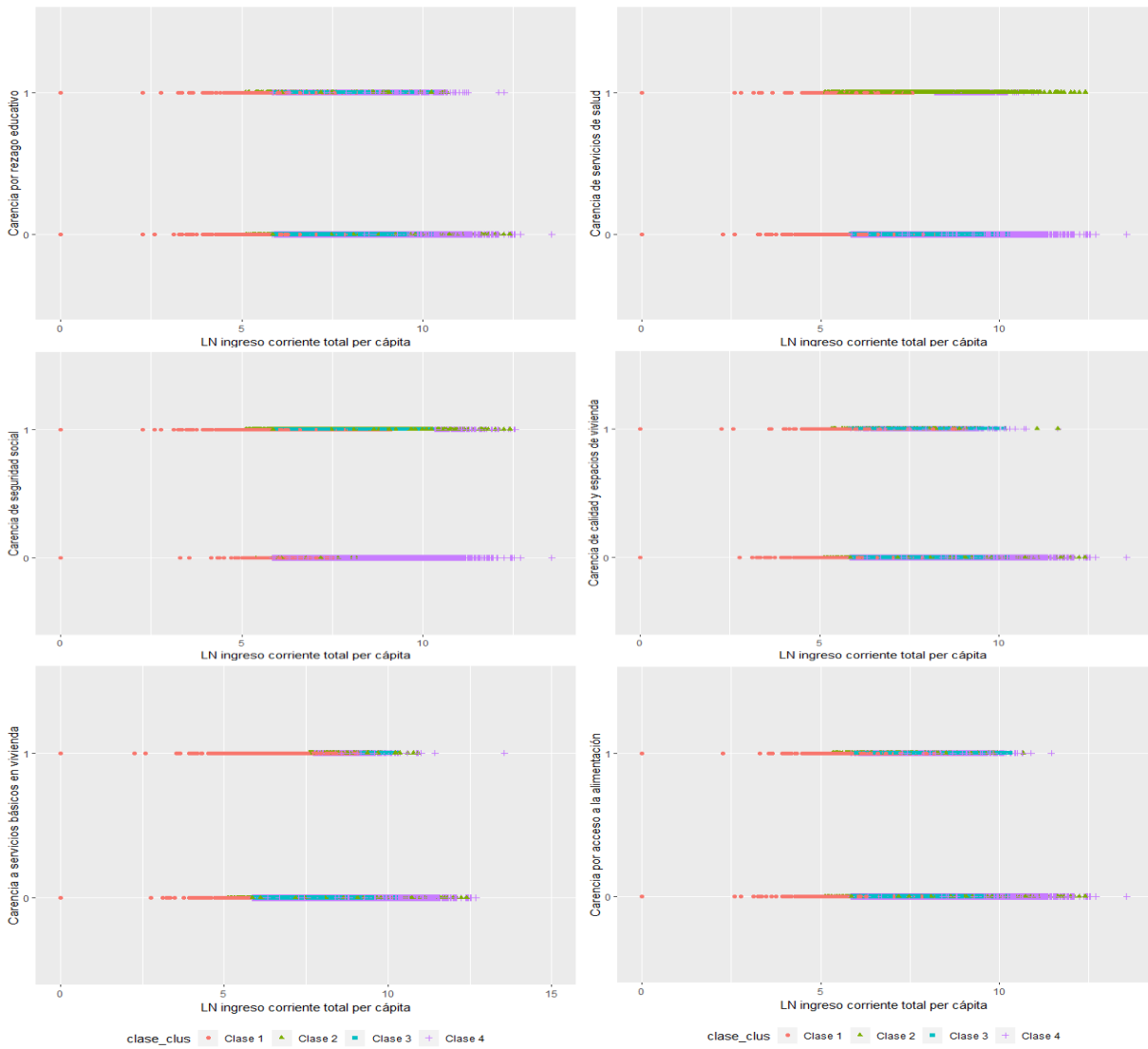
La **clase 4** contiene individuos en todas las clases respecto a la clasificación del CONEVAL, en 2016 y 2018, pero se diferencia de otras clases en que sus elementos presentan niveles reducidos respecto a los diferentes indicadores de carencia y los niveles de ingreso son más altos.

Clase	Indicador de carencia por rezago educativo	Indicador de carencia por acceso a los servicios de salud	Indicador de carencia por acceso a la seguridad social	Indicador de carencia por calidad y espacios de la vivienda	Indicador de carencia por acceso a los servicios básicos en la vivienda	Indicador de carencia por acceso a la alimentación
<b>Clase 1</b>	31%	9%	88%	29%	98%	35%
<b>Clase 2</b>	18%	100%	98%	9%	11%	20%
<b>Clase 3</b>	16%	0%	100%	10%	0%	20%
<b>Clase 4</b>	14%	1%	0%	5%	6%	15%

**Cuadro 15.** *Proporción de carencias presentadas respecto del total de población en cada clúster, 2018.*

Fuente: Elaboración propia con información del CONEVAL.





**Gráfico 15.** *Diferencias identificadas en los grupos por el contraste de carencias de derechos y el ingreso, 2018.*

Fuente: Elaboración propia con información del CONEVAL.

### 3.2.4 Resultados comparados

En los cuadros<sup>19</sup> 16 y 17 se presenta un resumen de los resultados de los algoritmos de clasificación para los cuales se pueden comparar directamente los resultados y verificar las diferencias<sup>20</sup> lo cual significa una ventaja. Sin embargo esto también puede significar una desventaja, ya que en el proceso de entrenamiento del algoritmo se usaron los resultados de clasificación del CONEVAL, lo que implica un sesgo favorable para estos resultados.

Así, si se asume que la clase 1 corresponde a los pobres, la clase 2 a los vulnerables por carencias, la clase 3 a los vulnerables por ingresos y la clase 4 a los no pobres se observan los siguientes resultados:

- Para 2016, se percibe que comparados con los resultados CONEVAL tanto en la Regresión Logística como en Naïve Bayes; existe una disminución de los casos clasificados como no vulnerables por ingresos o carencias y su vez se incrementan los casos de pobres y no pobres ni vulnerables.
- Para 2018, se observa un comportamiento distinto en los resultados de Naïve Bayes; son menores los clasificados como pobres y vulnerables por ingresos y mayores los clasificados como vulnerables por carencias y no pobres ni vulnerables respecto de los resultados del CONEVAL.

Categoría de Pobreza	CONEVAL	Regresión Logística Multinomial	Naïve Bayes
Pobreza	105,766	108,368	106,844
Vulnerable por carencias	81,349	78,747	80,695
Vulnerable por ingresos	16,261	16,210	13,379
No pobre ni vulnerable	54,271	54,322	56,729
Total	257,647	257,647	257,647

**Cuadro 16.** Comparación de resultados con algoritmos de aprendizaje supervisado, 2016  
Fuente: Elaboración propia con información del CONEVAL.

<sup>19</sup> Los datos presentados en los cuadros no consideran los registros que se encuentran nulos.

<sup>20</sup> Es importante mencionar que estas comparaciones no consideran la expansión con la variable factor, que se mencionó en el Capítulo 1, lo anterior con el objeto de identificar únicamente la diferencias de entre los datos utilizados.

<b>Categoría de Pobreza</b>	<b>CONEVAL</b>	<b>Regresión Logística Multinomial</b>	<b>Naïve Bayes</b>
Pobreza	107,558	110,234	106,461
Vulnerable por carencias	89,949	87,273	91,427
Vulnerable por ingresos	15,974	15,799	12,921
No pobre ni vulnerable	55,575	55,750	58,247
Total	269,056	269,056	269,056

**Cuadro 17.** Comparación de resultados con algoritmos de aprendizaje supervisado, 2018  
Fuente: Elaboración propia con información del CONEVAL.

En lo que respecta a los resultados del algoritmo de agrupamiento, es importante señalar que no se pueden asociar directamente sus resultados con los del CONEVAL, ya que no existe un entrenamiento a partir de dicha información y se hacen los grupos de acuerdo a las similitudes entre las variables usadas. No obstante, una potencial ventaja de esta técnica es que permite observar resultados que no se asocian a la metodología del CONEVAL y sí se asocian directamente al comportamiento de los datos lo que lleva a la posibilidad de observar resultados muy distintos y por lo mismo interesantes.

Al comparar los resultados entre los datos que se publicaron para 2016 y para 2018 (cuadro 18) se observan resultados distintos, aunque los resultados se mantienen en la Clase 2 y la Clase 4. Por otro lado, una particularidad que se debe considerar en los resultados es que éstos pueden variar en cada ejecución debido al componente aleatorio en el algoritmo.

De acuerdo con los resultados presentados, se observa que la Clase 4 agrupó al mayor número de personas y como se había visto en los cuadros 12 y 13 contempla a todas las personas que originalmente fueron clasificadas como no pobres ni vulnerables, lo interesante es que también agrupó a algunas personas vulnerables por ingresos, por carencias y pobres, y la razón de este agrupamiento entre personas que, en primera instancia, consideramos distintas es que observan niveles de ingresos más altos y en el caso de personas vulnerables por carencias o pobres con índices de privación más pequeños.

<b>Resultados</b>	<b>Clase 1</b>	<b>Clase 2</b>	<b>Clase 3</b>	<b>Clase 4</b>
Resultados 2016	98,498	35,073	23,436	100,640
Resultados 2018	52,922	32,237	77,109	106,788

**Cuadro 18.** *Comparación de resultados aprendizaje no supervisado, 2016 y 2018*

Fuente: Elaboración propia con información del CONEVAL.



# Conclusiones

## Conclusiones

La medición de la pobreza no es una tarea sencilla porque requiere entender y generalizar una problemática social con diversidad de causas y comportamientos distintos bajo determinadas circunstancias. En este sentido, se debe señalar que el CONEVAL hace un esfuerzo importante por identificar las condiciones de vida que definen a una persona en condición de pobreza a partir de dos dimensiones que son: el bienestar que se cuantificó mediante el ingreso de la persona, y la dimensión de derechos que se determinó a través de la presencia o ausencia de la carencia de derechos básicos; y de esta forma buscó conjugar ambas dimensiones y robustecer la detección de pobreza en un ámbito multidimensional.

El análisis realizado considera el mismo insumo de datos que el CONEVAL construye e identifica como determinantes de la condición de pobreza, y de esta forma se asume la misma perspectiva respecto de sus causas y detonantes. No obstante, a partir de la aplicación de métodos descriptivos, así como la implementación de algoritmos de análisis de datos se logran identificar áreas de oportunidad que diferencian grupos de aquellos que aparecen en la metodología del CONEVAL, la cual se enfoca principalmente en el cumplimiento estricto de la doble condicionante en la dimensión de bienestar y de derechos. Así, las principales deducciones son las siguientes:

1. Las personas con niveles bajos de ingresos que superan la LPI por un margen pequeño y presenten un importante número de carencias (los cuales bajo la metodología de CONEVAL se consideran como vulnerables por carencias) son clasificadas por el algoritmo de regresión logística multinomial y Naïve Bayes en la misma clase que aquellos que el CONEVAL clasifica como pobres.
2. También en el caso de personas que no presentan carencias de derechos, pero tienen niveles bajos de ingresos, el algoritmo de Naïve Bayes las identifica en la misma clase de quienes cumplen con la doble condicionante. Y por otro lado este algoritmo también agrupa en una misma clase a las

personas con montos de ingresos altos y a aquellas con ninguna o muy pocas carencias de derechos, lo cual permite deducir que este algoritmo podría detectar que los ingresos altos pueden ayudar cuando se tienen carencias, y no precisamente se es vulnerable

3. En el caso del algoritmo de agrupamiento se observaron grupos muy diferentes a los de CONEVAL. Se percibe que los grupos tienden a diferenciarse por la combinación, intensidad y número de carencias presentadas, así como el nivel de ingresos.
4. Se logró apreciar que la metodología de CONEVAL puede presentar áreas de oportunidad en relación a que el criterio de doble condición para la identificación de pobreza es estricto, principalmente en hogares que actualmente son clasificados como vulnerables por carencias de derechos y que pueden presentar un importante número de ellas y además tener ingresos reducidos (es decir, superar por muy poco la LPI). La metodología actual pierde de vista esta situación, debido a que únicamente contempla que estas personas superan la línea de bienestar pero no sopesa todas sus carencias.
5. El usar como insumos para la implementación de los algoritmos los mismos datos que usa el CONEVAL, permite hacer una comparación más exacta de los resultados, así como una interpretación en la misma perspectiva, lo cual resulta una ventaja cuando se busca identificar áreas de oportunidad en la metodología de CONEVAL. Sin embargo, esta decisión puede disminuir el margen de acción de los algoritmos. Ya que, suministrando una mayor diversidad de datos a los algoritmos se podría construir una visión distinta del problema. Es importante mencionar que una limitante fue que únicamente se usaron 7 atributos de los cuales 6 eran categóricos binarios y uno numérico lo cual implicó usar algoritmos que pudieran dar resultados con la combinación de ambos tipos de datos.
6. Se considera que es necesario robustecer la metodología del CONEVAL, para que ésta sea más amplia, sensible y flexible para que pueda detectar cuando las personas se enfrentan a situaciones socioeconómicas complejas.

Así, de acuerdo con las experiencias internacionales revisadas para detección de pobreza en áreas donde los datos son limitados, en México podrían observarse mejoras y complementos a la metodología que actualmente se aplica para la detección de pobreza. Por ejemplo, además de incorporar mayor diversidad de datos se pueden hacer contrastes con los resultados obtenidos mediante imágenes satelitales, ya que en África se han localizado resultados interesantes observando tales técnicas.

Los resultados de la presente investigación no se han puesto directamente a disposición de investigadores del CONEVAL para su verificación y retroalimentación, lo cual es un área de oportunidad de la presente investigación. Es importante mencionar que a partir de los hallazgos encontrados emanan más preguntas que pudieran analizarse y probarse para mejorar el proceso de identificación de pobreza, lo cual es necesario realizar para robustecer y generar una propuesta interesante al CONEVAL.



## Bibliografía

- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer, pp. 380-381
- Blumenstock, J., Cadamuro, G., & Robert, O. (2015). *Predicting poverty and wealth from mobile phone metadata*, pp. 1073-1076. Recuperado el 28 de Mayo de 2020, de <https://science.sciencemag.org/content/350/6264/1073/tab-pdf>
- Burke, M., Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., . . . Ermon, S. (2020). *Using publicly available satellite imagery and deep learning to understand economic well-being in Africa*. *Nat Commun* 11, 2583. Recuperado el 28 de Mayo de 2020, de <https://www.nature.com/articles/s41467-020-16185-w>
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). (s.f). *MEDICIÓN DE LA POBREZA GLOSARIO*. Recuperado el 25 de Enero de 2020, de <https://www.coneval.org.mx/Medicion/Paginas/Glosario.aspx>
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). (2014). *Metodología para la medición multidimensional de la pobreza en México, segunda edición*. Recuperado el 25 de Enero de 2020, de <https://www.coneval.org.mx/Informes/Coordinacion/Publicaciones%20oficiales/MEDICION MULTIDIMENSIONAL SEGUNDA EDICION.pdf>
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). (2019). *Metodología para la medición multidimensional de la pobreza en México, tercera edición*. Recuperado el 25 de Enero de 2020, de <https://www.coneval.org.mx/InformesPublicaciones/InformesPublicaciones/Documents/Metodologia-medicion-multidimensional-3er-edicion.pdf>
- Hastie, T., Tibshirani, R., & Friedman, J.H. (2008). *The elements of statistical learning: data mining, inference, and prediction*. 2da ed. California: Springer, pp. 119-120, 485-487, 507.
- Huang, Z. (1997). *A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining*. Recuperado el 25 de Enero de 2020, de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.83&rep=rep1&type=pdf>
- Huang, Z. (1998). *Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values*, pp. 283-304. Recuperado el 25 de Enero de 2020, de

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.4028&rep=rep1&type=pdf>

Instituto Nacional de Estadística y Geografía (INEGI). (2017). *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). 2016 Nueva serie*. Recuperado el 28 de Mayo de 2020, de <https://www.inegi.org.mx/programas/enigh/nc/2016/>

Instituto Nacional de Estadística y Geografía (INEGI). (2019). *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). 2018 Nueva serie*. Recuperado el 28 de Mayo de 2020, de <https://www.inegi.org.mx/programas/enigh/nc/2018/>

Instituto Nacional de Estadística y Geografía (INEGI). (2019). *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). 2018 Nueva serie*, p. 4. Recuperado el 28 de Mayo de 2020, de [https://www.inegi.org.mx/contenidos/programas/enigh/nc/2018/doc/enigh18\\_descriptor\\_archivos\\_fd\\_ns.pdf](https://www.inegi.org.mx/contenidos/programas/enigh/nc/2018/doc/enigh18_descriptor_archivos_fd_ns.pdf)

Jean, N., Burke, M., Xie, M., Davis, M., Lobell, D. B., & Ermon, S. (2016). *Combining satellite imagery and machine learning to predict poverty*, pp. 790-794. Recuperado el 28 de Mayo de 2020, de <https://science.sciencemag.org/content/353/6301/790/tab-pdf>

Marín Diazaraque, J. (n.d.). *Transformaciones de variables*. (2004-2005). 1-5. Recuperado el 25 de Enero de 2020, de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/EDDescrip/tema4.pdf>

Naciones Unidas. (2009). *Rethinking Poverty: Report on the World Social Situation 2010*. Recuperado el 05 de Junio de 2020, de <https://www.un.org/esa/socdev/rwss/docs/2010/fullreport.pdf>

Programa de las Naciones Unidas para el Desarrollo (PNUD). (2019). *The 2019 Global Multidimensional Poverty Index (MPI)*. Recuperado el 05 de Junio de 2020, de <http://hdr.undp.org/en/2018-MPI>

Ramasubramanian, K., & Singh, A. (2019). *Machine Learning Using R*. 2da ed. (s.l.): Apress, pp. 326-329, 376-385

Rojas, M. (2014). *Estimación de escalas de equivalencia en México*, pp. 4-17. Recuperado el 25 de Enero de 2020, de [https://www.inegi.org.mx/rde/rde\\_13/doctos/rde\\_13\\_art1.pdf](https://www.inegi.org.mx/rde/rde_13/doctos/rde_13_art1.pdf)

United Nations Economic Commission For Europe (UNECE). (2017). *Guide on Poverty Measurement*, p. 90. Recuperado el 25 de Enero de 2020, de

[https://ec.europa.eu/eurostat/ramon/statmanuals/files/UNECE\\_Guide\\_on\\_Poverty\\_Measurement.pdf](https://ec.europa.eu/eurostat/ramon/statmanuals/files/UNECE_Guide_on_Poverty_Measurement.pdf)

# ANEXOS

## ANEXO 1 Referencias particulares sobre los algoritmos implementados

	Referencias conceptuales consultadas	Referencias para la ejecución del algoritmo
<b>Regresión Logística Multinomial</b>	<p>Hastie, T., Tibshirani, R., &amp; Friedman, J.H. (2008). <i>The elements of statistical learning: data mining, inference, and prediction</i>. 2da ed. California: Springer, pp. 119-120.</p> <p>Ramasubramanian, K., &amp; Singh, A. (2019). <i>Machine Learning Using R</i>. 2da ed. (s.l.): Apress, pp. 326-329.</p>	<p>Se usó la librería nnet de R mediante la aplicación de <i>mmultinom</i>.</p> <p><i>Multinom</i>. (s.f.). Recuperado el 02 de Junio de 2020, de <a href="https://www.rdocumentation.org/packages/nnet/versions/7.3-14/topics/multinom">https://www.rdocumentation.org/packages/nnet/versions/7.3-14/topics/multinom</a></p>
<b>Naïve Bayes</b>	<p>Bishop, C. M. (2006). <i>Pattern recognition and machine learning</i>. New York: Springer, pp. 380-381.</p> <p>Ramasubramanian, K., &amp; Singh, A. (2019). <i>Machine Learning Using R</i>. 2da ed. (s.l.): Apress, pp. 376-385.</p>	<p>Se usó la librería e1071 de R mediante la aplicación de <i>naiveBayes</i>.</p> <p><i>naiveBayes</i>. (s.f.). Recuperado el 02 de Junio de 2020, de <a href="https://www.rdocumentation.org/packages/e1071/versions/1.7-3/topics/naiveBayes">https://www.rdocumentation.org/packages/e1071/versions/1.7-3/topics/naiveBayes</a></p>

<p><b>K-prototypes</b></p>	<p>Hastie, T., Tibshirani, R., &amp; Friedman, J.H. (2008). <i>The elements of statistical learning: data mining, inference, and prediction</i>. 2da ed. California: Springer, pp. 485-487, 507.</p> <p>Huang, Z. (1997). <i>A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining</i>. Recuperado el 25 de Enero de 2020, de <a href="http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.83&amp;rep=rep1&amp;type=pdf">http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.83&amp;rep=rep1&amp;type=pdf</a></p> <p>Huang, Z. (1998). <i>Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values</i>, pp. 283-304. Recuperado el 25 de Enero de 2020, de <a href="http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.4028&amp;rep=rep1&amp;type=pdf">http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.4028&amp;rep=rep1&amp;type=pdf</a></p>	<p>Se usó la librería clustMixType de R mediante la aplicación de <i>k-proto</i>.</p> <p><i>k-proto</i>. (s.f.). Recuperado el 02 de Junio de 2020, de <a href="https://www.rdocumentation.org/packages/clustMixType/versions/0.1-16/topics/kproto">https://www.rdocumentation.org/packages/clustMixType/versions/0.1-16/topics/kproto</a></p>
----------------------------	---	--

## Índice de términos

### “A”

Agrupamiento (Clustering).....	40
Aprendizaje no supervisado .....	40
Aprendizaje supervisado .....	31

### “C”

CONEVAL .....	1, 10
---------------	-------

### “D”

Doble condición.....	1, 8
----------------------	------

### “E”

ENIGH .....	2, 10
Espacio de bienestar .....	11
Espacio de derechos.....	13

### “I”

ICTPC.....	11,12
Índice de privación .....	15
Intensidad de la pobreza .....	17
IPM.....	5

### “K”

K-means .....	40, 41
K-modes .....	41
K-prototypes .....	41

### “L”

LPI.....	12
Línea de pobreza extrema por ingresos .....	29

### “M”

Matriz de confusión .....	33
Medidas de incidencia.....	16
Metodología de medición multidimensional de la pobreza...	11

### “N”

Naïve Bayes .....	35
No pobre multidimensional y no vulnerable .....	16

### “P”

Pobreza extrema .....	29
Pobres multidimensionales .....	15
Pobreza multidimensional .....	15
Probabilidad condicional .....	35
Profundidad de la pobreza .....	16

### “R”

Regresión logística.....	31
Regresión logística multinomial.....	31

### “V”

Vulnerables por carencias sociales	15
Vulnerables por ingresos.....	15

