

Reconocimiento estadístico de patrones - Proyecto 01
Índice de marginación en México dentro del periodo 2015 - 2020
Giovanni Gamaliel López Padilla

1. Introducción

Uno de los problemas que se presentan en las sociedades humanas, más en específico en la sociedad mexicana es el difícil acceso a los servicios públicos como salud, educación o justicia. Causando condiciones de vulnerabilidad e inseguridad. Una consecuencia de esta precariedad es la disminución en la calidad de vida de los habitantes. Con este problema surgen instituciones nacionales e internacionales para estudiar el caso, recabar información y generar indicadores que ayuden a identificar los casos en una región extensa. El consejo Nacional de Población (CONAPO) calcula el índice de marginación (IM) basado en el fundamento en el enfoque de necesidades básicas insatisfechas, utilizando el análisis de componentes principales¹. De una manera semejante, el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) estima el índice de rezago social (IRS) por medio de una suma ponderada de porcentajes². El IM se utiliza para localizar las zonas de atención prioritaria. Estas zonas tienen acceso a fondos y programas gubernamentales para el combate de la marginación y el desarrollo de la población³⁻¹¹. Los programas que usan al IM para determinar las localidades que serán beneficiarias son:

- Estrategia 100x100¹².
- Estrategia del Gobierno Federal para la Dotación de Piso Firme¹³.
- Programa de Apoyo Alimentario en Zonas de Atención Prioritaria¹⁴.
- Programa de Apoyo a Zonas de Atención Prioritaria¹⁵.

El IM utiliza como indicadores¹⁶ los siguientes porcentajes:

- Población analfabeta de 15 años o más.
- Población de 15 años o más sin educación básica.
- Viviendas particulares sin agua entubada.
- Viviendas particulares sin drenaje ni excusado.
- Viviendas particulares sin energía eléctrica.
- Viviendas particulares con piso de tierra.
- Viviendas particulares con hacinamiento
- Porcentaje de población que vive en localidades menores a 5000 habitantes.
- Población ocupada con ingresos de hasta 2 salarios mínimos.

Para estimar los índices de marginación se utiliza el método de distancia P_2^{17} (DP_2). El método de distancia P_2 fue desarrollado para hacer comparaciones temporales y espaciales. Ha sido aplciado en investigaciones sobre la calidad de vida en la Unión Europea y España¹⁸⁻²¹. La base del método es tener una base de referencia. Los valores de referencia pueden conformarse por los mínimos o máximos para todas los indicadores o para cada indicador. Estos valores no tienen que ser necesariamente reales. Si se toman los valores mínimos, entonces el indicador de DP_2 tiene valores altos cuando existen mejores condiciones socioeconomicas. La base de referencia a nivel municipal para el periodo 2010-2020 se muestra en la tabla 1¹⁶. El DP_2 evita la duplicación de información y imparcialidad en el esquema de ponderación. Además de esto contiene las siguientes propiedades matemáticas: existencia, determinación monotonía, unicidad, invarianza frente a la base de referencia, homogeneidad, transitividad, exhaustiva, aditividad y neutralidad^{18;22}. Debido a las propiedades presentes en el indicador DP_2 dan un beneficio al análisis de componentes principales¹⁸.

Indicadores socioeconomicos	Base de referencia
Porcentaje de población analfabeta de 15 años o más.	-66.74
Porcentaje de población de 15 años o más sin educación básica.	-94.79
Porcentaje de viviendas particulares sin agua entubada.	-89.90
Porcentaje de viviendas particulares sin drenaje ni excusado.	-69.45
Porcentaje de viviendas particulares sin energía eléctrica.	-99.74
Porcentaje de viviendas particulares con piso de tierra.	-79.71
Porcentaje de viviendas particulares con hacinamiento	-83.24
Porcentaje de porcentaje de población que vive en localidades menores a 5000 habitantes.	-100.00
Porcentaje de población ocupada con ingresos de hasta 2 salarios mínimos.	-100.00

Tabla 1: Base de referencia a nivel municipal en el periodo 2010-2020.

El indicador DP_2 sintético se define en la ecuación 1.

$$DP_2 = \sum_i \frac{d_{ij}}{\sigma_j} (1 - R_{j,j-1,\dots,1}^2) \quad R_1^2 = 0 \quad (1)$$

donde d_{ij} es la distancia de la j -ésima variable del municipio con respecto a la base de referencia, σ_j es la desviación estándar de la variable j y $R_{j,j-1,\dots,1}^2$ es el coeficiente de determinación de la regresión del indicador parcial j con los demás indicadores. R_1^2 es igual a cero porque el primer indicador aporta toda la información y al no existir algún indicador previo su ponderación es la unidad. Una forma de observar el índice de marginación es normalizando (ecuación 2) sus valores para apreciar la evolución de cada municipio²³.

$$\bar{DP}_2^i = \frac{DP_2^i - \min(DP_2)}{\max(DP_2) - \min(DP_2)} \quad (2)$$

Es importante aclarar que el índice de marginación normalizado se puede tomar como una guía ya que es sensible a los valores atípicos, lo que en cierta medida hace que se amplifiquen los datos normalizados ante los casos más extremos. Pero debido a las propiedades del método y que el rango de cada uno de los índices es estrecho y lineal, se podría esperar que la normalización no genere ruido en las colas de la distribución. Cada indicador es normalización usando la ecuación 3.

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (3)$$

Donde X_{ij} representa al i -ésimo valor de la variable k , μ_k y σ_k es el promedio y desviación estandar de la variable k respectivamente. Debido a la normalización de cada indicador, el

IM no puede utilizarse para realizar un seguimiento temporal. Para representar a los nueva indicadores en uno solo se aplica el análisis de componentes principales. La CONAPO emplea el método de Dalenius Hodges²⁴ para agrupar los puntajes en cinco categorías de marginación (muy alta, alta, media, baja y muy baja). El IM permite ordenar y clasificar según su categoría de marginación en el año que se calcula.

2. Métodos y materiales

2.1. Algoritmos de reducción de dimensionalidad

La reducción de dimensionalidad es un proceso que disminuye el tamaño de variables o que describen a un conjunto de datos. Las ventajas de la reducción de dimensionalidad son las siguientes:

- Reduce el espacio de tiempo y almacenamiento requerido.
- La eliminación de multicolinealidad mejora el rendimiento del modelo de aprendizaje automático.
- Se hace más fácil de visualizar los datos cuando se reduce dos o tres dimensiones.

Los algoritmos que se implementaron en este trabajo fueron ISOMAP, K-means, Self-Organizing-Map (SOM), T-SNE y análisis de componentes principales (PCA).

2.1.1. ISOMAP

El método ISOMAP esta basado en la teoría espectral, la cual trata de preservar la distancia euclidiana en la reducción de dimensionalidad a la cual se quiere llegar. El método empieza creando un grafo de vecinos. En seguida de calcula la geodesica entre cada par de elementos de la base de datos, creando así una matriz de disimilaridad. A partir de la matriz de disimilaridad se emplea el escalamiento multidimensional.

2.1.2. K-means

El algoritmo de K-means es un algoritmo de clasificación no supervisada que agrupa objetos en k conjuntos basandose en los valores de las variables dadas. El algoritmo asigna a cada elemento a un conjunto minimizando la suma de las distancias de cada objeto con el centroide del conjunto (ecuación 4).

$$\min_{\{r_{nk}\}\{\mu_k\}} \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2 \quad (4)$$

2.1.3. Self-Organizing-Map (SOM)

SOM es un algoritmo de clasificación no supervisada. Su objetivo es reducir la dimensionalidad del los datos dados preservando la estructura topológica. El algoritmo minimiza la ecuación 5.

$$C(\{m_l\}, \{l(i)\}) = \sum_i \sum_k h(\|c_k - c_{l(i)}\|^2) \|x_i - m_k\|^2 \quad (5)$$

Donde m_l es el representante más cercano a x_i , $h()$ es una función decreciente y positiva y c_k son las celdas donde se asociara el dado x_i .

2.1.4. T-SNE

El algoritmo de T-SNE consiste en crear una distribución de probabilidad que represente las similitudes entre vecinos en un espacio de gran dimensión en un espacio de menor dimensión. Para cada elemento del conjunto de datos se centra una distribución gaussiana alrededor del elemento. En seguida se obtiene la densidad bajo la distribución y normalizamos el valor, calculando así una probabilidad condicional (ecuación 6).

$$P_i = P_{j|i} = \frac{\exp(-||x_i - x_j||^2/\sigma)}{\sum_{k \neq i} \exp(-||x_k - x_i||^2/\sigma)} \quad (6)$$

El valor de σ se define por medio de un parámetro llamado perplejidad, el cual corresponde al número de vecinos alrededor de cada punto. A una mayor probabilidad condicional, los elementos ij son más similares. A partir de las probabilidades condicionales se tratan encontrar pares de elementos tales que las distribuciones $P_{i|j}$ y $q_{i|j}$ se parezcan. El valor de $q_{i|j}$ se obtiene mediante la distribución T (ecuación 7).

$$Q_i = q_{j|i} = \frac{\exp(-||x_i - x_j||^2)}{\sum_{k \neq i} \exp(-||x_k - x_i||^2)} \quad (7)$$

Por lo tanto, el algoritmo T-SNE se centra en minimizar la distancia Kullback-Leiber (ecuación 8)

$$\min d_i(P_i; Q_i) \quad d(P^1; P^2) = \sum_i P_i^1 \log \left(\frac{P_i^1}{P_i^2} \right) \quad (8)$$

2.1.5. Análisis de componentes principales

El análisis de componentes principales (PCA) es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información. La idea del algoritmo es asociar a cada elemento de la base de datos con un vector de menor dimensión tal que se minimiza la ecuación 9.

$$\sum_i \sum_j (d(x_i, x_j)^2 - d(x_i^* - x_j^*)^2)^2 \quad (9)$$

El problema de minimizar la ecuación 9 se puede transformar a la ecuación 10.

$$\min || -\frac{1}{2}C(D^2 - D^{*2})C ||_F^2 \quad (10)$$

Donde C es la matriz para centrar. Definiendo a la matriz kernel como $K = XX^t$ y D como la matriz asociada a las distancias, la ecuación 10 puede escribirse como en la ecuación 11.

$$\min ||K_c - K^*|| \quad K_c = CKC \quad (11)$$

La solución de la ecuación 11 es $K = \sum_i^p \lambda_i v_i v_i^t$, donde v_i y λ_i son los eigenvectores y eigenvalores de K respectivamente.

2.2. Base de datos de CONAPO

Se descargo la base de datos de índice de marginación por municipio del periodo 1990 a 2020 disponibles desde el portal de la Consejo Nacional de Población (CONAPO)^{25;26}. Los datos que contienen los archivos se encuentran descritos en la tabla 2.

Abreviación	Descripción
CVE_ENT	Clave de entidad federativa
NOM_ENT	Nombre de entidad federativa
CVE_MUN	Clave del municipio
NOM_MUN	Nombre del municipio
POB_TOT	Población total
ANALF	Porcentaje de población analfabeta de 15 años o más
SBASC	Porcentaje de población de 15 años o más sin educación básica
OVSDE	Porcentaje de ocupantes en viviendas particulares habitadas sin drenaje ni excusado
OVSEE	Porcentaje de ocupantes en viviendas particulares habitadas sin energía eléctrica
OVSDE	Porcentaje de ocupantes en viviendas particulares habitadas sin agua entubada
OVPT	Porcentaje de ocupantes en viviendas particulares habitadas con piso de tierra
VHAC	Porcentaje de viviendas particulares con hacinamiento
PL.5000	Porcentaje de población que vive en localidades menores a 5 000 habitantes
PO2SM	Porcentaje de población ocupada con ingresos de hasta 2 salarios mínimos
IM	Índice de marginación
GM	Grado de marginación

Tabla 2: Diccionario de los indicadores socioeconomicos e índice de marginación por municipio.

Se realizaron diversas modificaciones a los datos obtenidos de CONAPO. Las modificaciones que se realizaron son las siguientes:

- Para el archivo que contiene los datos del año 2015 se elimino la linea 1555 debido a que no se recabaron datos. Los datos correspondiente a esa linea son para el municipio de Nicolás Ruíz, Chiapas.
- Las celdas que contenian el símbolo - fueron eliminadas. Esto es porque el símbolo fue utilizado para señalar que no se recabo el datos correspondiente.
- Los nombres de las cabeceras de columna fue modificado para que fueran igual a las mostradas en la tabla 2.
- Del archivo que contiene los datos del año 2020 se produjo un archivo auxiliar del tipo *csv*. En el archivo auxiliar se introdujo la información que se encuentra en la hoja llamada *IMM.2020*.

En la figura 1 se muestra la frecuencia relativa de los índices de marginación para los 2469 municipios del año 2015 y 2020.

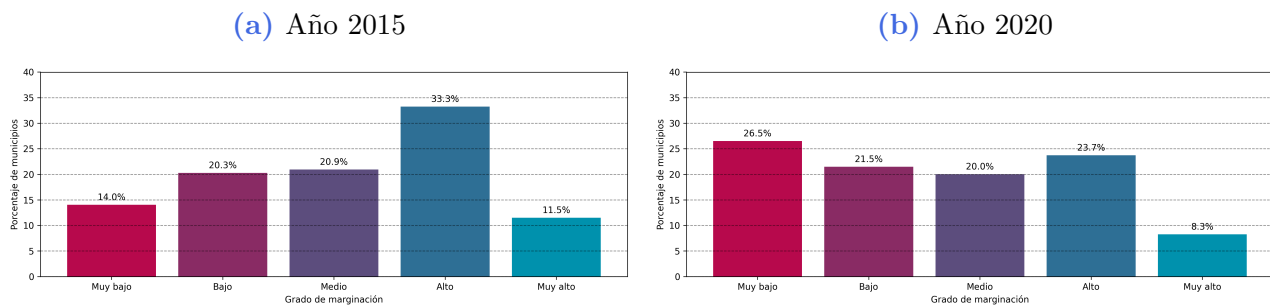


Figura 1: Frecuencia relativa de cada índice de marginación en los años 2015-2020 a nivel municipio.

En la figura 2 se muestran los índices de marginación para cada municipio en los años 2015 y 2020. Los datos para delimitar cada municipio fue obtenido del siguiente [blog](#)²⁷. La distribución de los índices de marginación en la República Mexicana marca que se llegan a obtener valores bajos en la zona norte y valores altos en la zona del centro y sur del país.

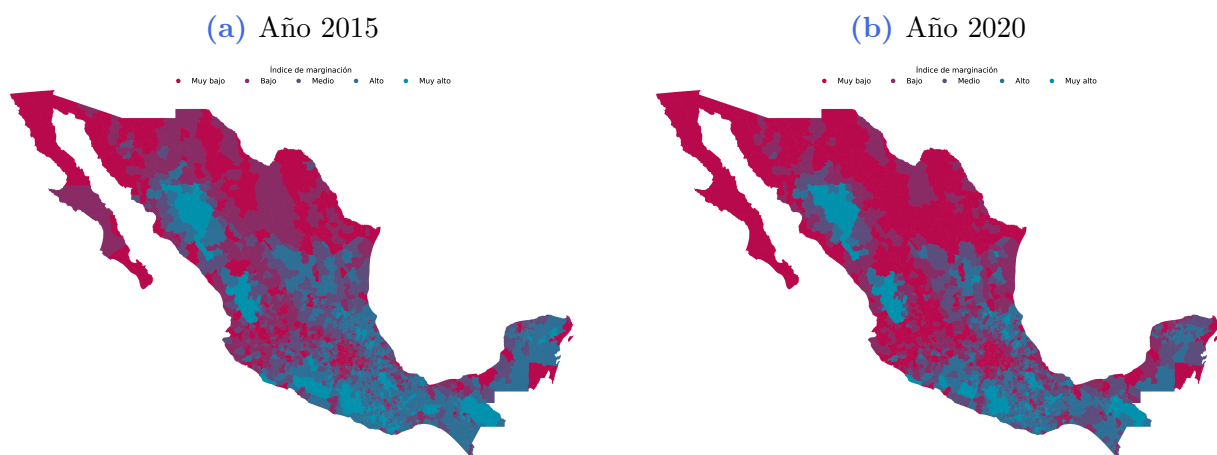


Figura 2: Grado de marginación a nivel municipio en la República Mexicana.

A partir de las columnas de cada indicador (ANALF, SBASC, OVSDE, OVSEE, OVSAE, OVPT, VHAC, PL.50 y PO2SM) se creó una matriz de embeddings a la cual se normalizó siguiendo la ecuación 3. Esta matriz será la que se le dará como valor de entrada a cada algoritmo presentado en la sección 2.1.

3. Resultados

4. Conclusiones

5. Referencias

- [1] Consejo Nacional de Población (CONAPO). Índices de marginación 2020; 2021. Available from: https://www.gob.mx/cms/uploads/attachment/file/685354/Nota_te_cnica_IMEyM_2020.pdf.
- [2] Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). Los mapas de Pobreza en México. Anexo técnico metodológico; 2007. Available from: https://www.coneval.org.mx/rw/resource/coneval/med_pobreza/1024.pdf.

- [3] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2012.; 2011. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2011_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2012.pdf.
- [4] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2013.; 2012. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2012_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2013.pdf.
- [5] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2014.; 2013. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2013_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2014.pdf.
- [6] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2015.; 2014. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2014_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2015.pdf.
- [7] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2016.; 2015. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2015_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2016.pdf.
- [8] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2017.; 2016. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2016_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2017.pdf.
- [9] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2018.; 2017. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2017_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2018.pdf.
- [10] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2019.; 2018. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2018_BIENESTAR_Declaratoria_zonas_atencion_prioritaria_2019.pdf.
- [11] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2020.; 2019. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2019_BIENESTAR_Declaratoria_zonas_atencion_prioritarias_2020_20191211.pdf.
- [12] Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). Evaluación de impacto de la Estrategia 100x100; 2013. Available from: https://www.coneval.org.mx/Informes/Evaluacion/Impacto/Evaluacion_de_impacto_de_la_Estrategia_100x100.pdf.
- [13] Diario Oficial de la Federación (DOF). Reglas de Operación del Programa de Vivienda Social.; 2020. Available from: https://www.dof.gob.mx/nota_detalle.php?codigo=5585435&fecha=04/02/2020.
- [14] Subsecretaría de Desarrollo Social y Humano (SEDESOL). Análisis de funcionamiento y operación del programa de apoyo alimentario en zonas de atención prioritaria; 2008. Available from: https://www.coneval.org.mx/Informes/Evaluacion/Complementarias/Complementarias_2008/SEDESOL/compl_08_sedesol_PAAZAP.pdf.

- [15] Diario Oficial de la Federación (DOF). Acuerdo por el que se emiten las Reglas de Operación del Programa para el Desarrollo de Zonas Prioritarias, para el ejercicio fiscal 2015; 2014. Available from: https://www.gob.mx/cms/uploads/attachment/file/26103/rop_pdzp.pdf.
- [16] Consejo Nacional de Población (CONAPO). Índices de marginación; 2021. Available from: https://www.gob.mx/cms/uploads/attachment/file/685354/Nota_te_cnica_IMEyM_2020.pdf.
- [17] Trapero XBP. Problemas de la medición del bienestar y conceptos afines: (una aplicación al caso español). Presidencia del Gobierno, Instituto Nacional de Estadística; 1977.
- [18] Somarriba N, Pena B. Synthetic Indicators of Quality of Life in Europe. Social Indicators Research. 2008 nov;94(1):115–133. Available from: <https://doi.org/10.1007%2Fs11205-008-9356-y>.
- [19] Zarzosa P. The social welfare in Spain before the crisis: Territorial and chronological analysis. International Journal of Advances in Management and Economics. 2012;1(4):165–171.
- [20] Espina Z, del Pilar M. Aproximación a la medición del bienestar social. Idoniedad del indicador sintético”Distancia-P2”. Consejo Superior de Investigaciones Científicas. 1996;.
- [21] Nayak P, Mishra SK. Efficiency of Pena’s P2 Distance in Construction of Human Development Indices. SSRN Journal Electronic Journal. 2012;Available from: <https://doi.org/10.2139%2Fssrn.2066567>.
- [22] Espina PZ, Arechavala NS. An Assessment of Social Welfare in Spain: Territorial Analysis Using a Synthetic Welfare Indicator. Soc Indic Res. 2012 feb;111(1):1–23. Available from: <https://doi.org/10.1007%2Fs11205-012-0005-0>.
- [23] Somorriba Arechavala N, Zarzosa Espina P, Pena Trapero B. La calidad de vida en la unión europea. Un análisis temporal por medio de indicadores sintéticos; 2013. Available from: <https://old.reunionesdeestudiosregionales.org/Oviedo2013/htdocs/pdf/p851.pdf>.
- [24] Dalenius T, Hodges JL. Minimum Variance Stratification. Journal of the American Statistical Association. 1959;54(285):88–101. Available from: <http://www.jstor.org/stable/2282141>.
- [25] Consejo Nacional de Población (CONAPO). Índice de marginación por municipio, 1990-2015;. Available from: http://www.conapo.gob.mx/work/models/CONAPO/Marginacion/Datos_Abiertos/Municipio/Base_Indice_de_marginacion_municipal_90-15.csv.
- [26] Consejo Nacional de Población (CONAPO). Índice de marginación por municipio 2020;. Available from: http://www.conapo.gob.mx/work/models/CONAPO/Marginacion/Datos_Abiertos/Municipio/IMM_2020.xls.
- [27] Santos Ochoa JJ. Delimitaciones de los municipios de México; 2021. Available from: http://jjsantoso.com/blog/datos/mapa_mexico.zip.