



# Un modelo de la clase media mexicana a partir de redes neuronales.

[corderopablo94@gmail.com](mailto:corderopablo94@gmail.com)

H. Reyes Cervantes, H. Vaquera Huerta y P. Castillo Cordero.  
FCFM-BUAP



## Introducción.

Este trabajo consistió en la aplicación de la metodología de redes neuronales en la predicción de que un hogar pertenezca a la clase media. Para ello se tomó la encuesta nacional de ingresos y gastos de los hogares hecha en 2016 por el Instituto Nacional de Estadística, Geografía e Informática (INEGI).

## Caso de estudio.

La encuesta nacional de ingresos y gastos de los hogares (ENIGH) es un instrumento desarrollado por el INEGI. Dicha encuesta tiene una periodicidad de 2 años y sirve para brindar un panorama de los ingresos la distribución en los hogares mexicanos.

La encuesta realizada en el año 2016, consta de una muestra de tamaño N=70311 viviendas distribuidas en las 32 entidades federativas.

Para la definición de la clase media se utilizará la que nos brinda la OCDE: "La clase media son aquellas personas que reciben ingresos de entre 75% y 200% del ingreso nacional mediano" [1].

## VARIABLES DE INTERÉS.

Se optó por escoger aquellas variables que proporcionan mayor información sobre un hogar y las personas que lo habitan así como la variable de ingreso del hogar. Teniendo un total de 17 variables las cuales son:

Tamaño de localidad, estrato socioeconómico, clase de hogar, sexo del jefe, edad del jefe, número de integrantes del hogar, número de integrantes del hogar hombres, número de integrantes del hogar mujeres, mayores de 12 años, menores, integrantes de 12 a 64 años, y 65 y más, ocupados, perceptores de ingreso, perceptores de ingreso ocupados, educación formal. Estos datos fueron utilizados sin modificaciones salvo en el caso de la educación, donde se reorganizaron los posibles valores de esta variable categórica y en el caso de ingreso se utilizaron los datos numéricos para convertirla en una variable dummy donde el caso de éxito es pertenecer al menos a la clase media.

## ANÁLISIS DE LOS DATOS CON REDES NEURONALES

Para el análisis de los datos nos apoyaremos del software SPSS y se utilizará el procedimiento de perceptrón multicapa (MLP) utilizando los parámetros predefinidos por el software [2].

La estructura del procedimiento MLP está compuesta por:

- La capa de entrada, la cual contiene a los predictores.
- Las capas ocultas que contienen los nodos no observables
- La capa de salida, la cual contiene las respuestas [3].

## Resultados

Podemos observar en el resumen del procedimiento de los casos, que la muestra se dividió en 3 sub muestras: la de entrenamiento, la de prueba y la de reserva.

	N	Porcentaje
Muestra Entrenamiento	42076	59.8%
Prueba	21146	30.1%
Reserva	7089	10.1%
Válidos	70311	100.0%
Excluidos	0	
Total	70311	

## Resumen del modelo

Entrenamiento	Error de entropía cruzada	21659.053
	Porcentaje de pronósticos incorrectos	25.6%
	Regla de parada utilizada	1 pasos consecutivos sin disminución del error <sup>a</sup>
	Tiempo de entrenamiento	00:00:10.545
Prueba	Error de entropía cruzada	10827.691
	Porcentaje de pronósticos incorrectos	25.3%
Reserva	Porcentaje de pronósticos incorrectos	26.9%

Variable dependiente: Ingreso corriente binario

a. Los cálculos del error se basan en la muestra de prueba.

Del resumen del modelo podemos destacar que el porcentaje de pronósticos incorrectos es del 26.9%.

## Clasificación

Muestra	Observado	Pronosticado		
		0	1	Porcentaje correcto
Entrenamiento	0	17798	4511	79.8%
	1	6257	13510	68.3%
	Porcentaje global	57.2%	42.8%	74.4%
Prueba	0	9016	2181	80.5%
	1	3166	6783	68.2%
	Porcentaje global	57.6%	42.4%	74.7%
Reserva	0	2947	790	78.9%
	1	1120	2232	66.6%
	Porcentaje global	57.4%	42.6%	73.1%

Variable dependiente: Ingreso corriente binario

Por último nos da un resumen de la clasificación obtenida a partir de nuestro modelo MLP, donde 1 significa que ese hogar pertenece al menos a la clase media bajo la definición de la OCDE. Cabe resaltar que los porcentajes globales de las particiones de la muestra rondan alrededor del 74% por lo cual podemos decir que nuestro modelo es un modelo suficientemente satisfactorio.

## Conclusión

Después de analizar el modelo arrojado podemos concluir que la selección de nuestras variables fue acertada ya que arroja porcentaje de pronósticos correctos suficientemente alto.

Este tipo de modelos son importantes para estudiar los movimientos de la economía de los hogares y adquieren una mayor dimensión en tiempos como los actuales donde los cambios en la sociedad son tan extremos. Dándonos una herramienta para poder comparar el estado de la economía familiar a través del tiempo.

## Referencias

[1] ¿Cuánto debes ganar en México para pertenecer a la clase media o alta? (2019, 11 abril). Forbes. <https://www.forbes.com.mx/cuanto-debes-ganar-en-mexico-para-pertenecer-a-la-clase-media-o-alta/>

[2] Hernández Martínez, A., & Reyes Cervantes, H. (2017). *Riesgo de Crédito: Análisis Mediante Redes Neuronales*. BUAP.

[3] Haykin S., *Neural Networks: A Comprehensive Foundation*, Prentice Hall, (1994).