

APRENDIZAJE NO SUPERVISADO PARA IDENTIFICAR FACTORES
SOCIOECONÓMICOS ASOCIADOS A DEFUNCIONES FETALES EN COLOMBIA

LICETH NATALÍ BAUTISTA RAMOS
JACKSYMAR PAOLA NUÑEZ ROSALES

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2025

APRENDIZAJE NO SUPERVISADO PARA IDENTIFICAR FACTORES
SOCIOECONÓMICOS ASOCIADOS A DEFUNCIONES FETALES EN COLOMBIA

LICETH NATALÍ BAUTISTA RAMOS
JACKSYMAR PAOLA NUÑEZ ROSALES

Trabajo de grado para optar al título de
Matemática

Director
Andrés Sebastián Ríos Gutiérrez
Candidato a Doctor

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2025

DEDICATORIA

Dedico este trabajo a mis padres, por haber creído siempre en mí; a la familia Peña Daza, que durante muchos años me apoyó en este recorrido; a Luis Eduardo, mi compañero de vida, quien me ha acompañado con su apoyo y amor incondicional; y a mi querido perrito Fray, cuya lealtad y cariño fueron un apoyo invaluable durante este proceso.

Liceth Natalí Bautista Ramos

A Dios, por guiar mis pasos y acompañarme en cada momento de dificultad y alegría; y a mi hermana Anyi, por su compañía, paciencia y aliento constante, que hicieron posible este logro.

Jacksymar Paola Nuñez Rosales

AGRADECIMIENTOS

Agradezco a los profesores Gilberto Arenas, Carlos Uzcátegui y Luis ángel por siempre apoyarme y tener fe en mí; al profesor Andrés, por haber guiado este trabajo y brindado consejos en momentos de frustración; y a los demás docentes de la carrera por sus valiosas enseñanzas.

También agradezco a mis amigos que me acompañaron al inicio de este largo camino: Julián y Natalia, quienes además de ofrecerme su dulce amistad, me ayudaron a comprender los temas cuando me resultaban difíciles; y a mi querida amiga incondicional Pilar, con quien atesoro los recuerdos que compartimos.

Finalmente, a mis compañeras con quienes culminé la carrera: Zareth, Yin y Jacksymar, por compartir momentos felices e inolvidables a mi lado.

Liceth Natalí Bautista Ramos

A mi mamá, por ser mi fuerza; a mi papá, por su apoyo; a mis hermanos, abuelos y a mi tía María, por darme equilibrio y recordarme siempre lo verdaderamente importante, y a mi familia que me acompañó siempre. A mis compañeros de universidad, por compartir este camino. A mi compañera de trabajo de grado, por impulsar este logro juntas. Finalmente, me agradezco a mí misma por no rendirme y seguir siempre adelante.

Jacksymar Paola Nuñez Rosales

CONTENIDO

	pág.
INTRODUCCIÓN	13
1 PLANTEAMIENTO	15
1.1 Descripción del problema	15
1.2 Justificación	15
1.3 Antecedentes	16
2 OBJETIVOS	18
2.1 Objetivo general	18
2.1.1 Objetivos específicos	18
3 FUNDAMENTOS DE REDES NEURONALES Y APRENDIZAJE NO SUPERVISADO	19
3.1 Conceptos básicos	19
3.2 Redes neuronales	19
3.3 Redes neuronales	34
3.4 Modelos no supervisados	35
4 ALGORITMOS A IMPLEMENTAR	38
4.1 Máquina de Boltzmann Restringida	38
4.1.0.1 Modelo gráfico y parametrización	39
4.2 El Algoritmo de K-Medias	41
4.3 Entropía	43

5 CASOS DE ESTUDIO: FACTORES SOCIOECONÓMICOS ASOCIADOS A MUERTES FETALES.	49
5.1 Descripción de la base de datos.	49
5.2 Análisis de factores socioeconómicos asociados a defunciones fetales.	55
5.2.1 Máquina de Boltzmann Restringida	56
5.2.1.1 RBM para la base de datos	56
5.2.1.2 Error de reconstrucción	60
5.2.2 Matriz de correlación	62
5.3 Etapas	65
5.3.1 Antes: K-Medias	66
5.3.2 Después: K-Medias	88
6 CONCLUSIONES	113
BIBLIOGRAFÍA	116
ANEXOS	122

LISTA DE FIGURAS

	pág.
Figura 1 Red neuronal Monocapa.	21
Figura 2 Representación matricial y vectorial de la red neuronal monocapa.	22
Figura 3 Red neuronal multicapa.	24
Figura 4 Función escalón	26
Figura 5 Función de activación lineal	27
Figura 6 Función sigmoide	27
Figura 7 Función Tangente Hiperbólica	28
Figura 8 Función ReLU	29
Figura 9 Función Leaky ReLU	29
Figura 10 Función ReLU	30
Figura 11 Función ELU	31
Figura 12 Función Swish	32
Figura 13 Función Softmax	32
Figura 14 Estructura de una RBM con dos nodos visibles y tres nodos ocultos.	39
Figura 15 Matriz de características latentes para la base de datos del año 2020.	58
Figura 16 Matriz de características latentes para la base de datos del año 2021.	59
Figura 17 Matriz de características latentes para la base de datos del año 2023.	60
Figura 18 Error de reconstrucción de la RBM para el año 2020.	61
Figura 19 Error de reconstrucción de la RBM para el año 2021.	61
Figura 20 Error de reconstrucción de la RBM para el año 2023.	61
Figura 21 Matriz de correlación después de RBM para el año 2020.	63
Figura 22 Matriz de correlación después de RBM para el año 2021.	64

Figura 23	Matriz de correlación después de RBM para el año 2023.	65
Figura 24	Agrupación por cluster para el año 2020.	67
Figura 25	Distribución de clústeres K-means vs Seguridad Social. Año 2020.	69
Figura 26	Distribución de clústeres K-means vs Edad Madre. Año 2020.	70
Figura 27	Distribución de clústeres K-means vs Edad Madre. Año 2020.	71
Figura 28	Distribución de clústeres K-means vs Código Departamento. Año 2020.	72
Figura 29	Entropía por clúster y variable socioeconómica. Año 2020.	74
Figura 30	Agrupación por cluster para el año 2021.	75
Figura 31	Distribución de clústeres K-means vs Seguridad Social. Año 2021.	77
Figura 32	Distribución de clústeres K-means vs Nivel Educativo. Año 2021.	78
Figura 33	Distribución de clústeres K-means vs Edad Madre. Año 2021.	78
Figura 34	Distribución de clústeres K-means vs Código Departamento. Año 2021.	79
Figura 35	Entropía por clúster y variable socioeconómica. Año 2021.	81
Figura 36	Agrupación por cluster para el año 2023.	82
Figura 37	Distribución de clústeres K-means vs Seguridad Social. Año 2023.	84
Figura 38	Distribución de clústeres K-means vs Nivel Educativo. Año 2023.	85
Figura 39	Distribución de clústeres K-means vs Edad Madre. Año 2023.	85
Figura 40	Distribución de clústeres K-means vs Código Departamento. Año 2023.	86
Figura 41	Entropía por clúster y variable socioeconómica. Año 2023.	88
Figura 42	Agrupación por cluster para el año 2020.	90
Figura 43	Distribución de clústeres K-means vs Seguridad Social. Año 2020.	93
Figura 44	Distribución de clústeres K-means vs Nivel Educativo. Año 2020.	93
Figura 45	Distribución de clústeres K-means vs Edad de la Madre. Año 2020.	94
Figura 46	Distribución de clústeres K-means vs Departamento de Residencia. Año 2020.	95
Figura 47	Entropía por clúster y variable socioeconómica. Año 2020.	97
Figura 48	Agrupación por cluster para el año 2021.	98

Figura 49	Distribución de clústeres K-means vs Seguridad Social. Año 2021.	100
Figura 50	Distribución de clústeres K-means vs Nivel Educativo. Año 2021.	101
Figura 51	Distribución de clústeres K-means vs Edad de la Madre. Año 2021.	101
Figura 52	Distribución de clústeres K-means vs Departamento. Año 2021.	102
Figura 53	Entropía por clúster y variable socioeconómica. Año 2021.	104
Figura 54	Agrupación por cluster para el año 2023.	105
Figura 55	Distribución de clústeres K-means vs Seguridad Social. Año 2023.	107
Figura 56	Distribución de clústeres K-means vs Nivel Educativo. Año 2023.	108
Figura 57	Distribución de clústeres K-means vs Edad de la Madre. Año 2023.	108
Figura 58	Distribución de clústeres K-means vs Departamento. Año 2023.	109
Figura 59	Entropía por clúster y variable socioeconómica. Año 2023.	110
Figura 60	Conjunto de datos	126
Figura 61	Aprendizaje Hebbiano	127
Figura 62	Funcionamiento de la red neuronal	128
Figura 63	Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2020.	131
Figura 64	Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2021.	131
Figura 65	Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2023.	132
Figura 66	Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2020.	132
Figura 67	Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2021.	133
Figura 68	Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2023.	133

LISTA DE TABLAS

	pág.	
Tabla 1	Notación empleada en la red neuronal.	24
Tabla 2	Descripción de las variables socioeconómicas, demográficas y médicas utilizadas en el análisis de defunciones fetales en Colombia.	54

RESUMEN

TÍTULO: APRENDIZAJE NO SUPERVISADO PARA IDENTIFICAR FACTORES SOCIOECONÓMICOS ASOCIADOS A DEFUNCIONES FETALES EN COLOMBIA *

AUTOR: LICETH NATALI BAUTISTA RAMOS - JACKSYMAR PAOLA NUÑEZ ROSALES **

PALABRAS CLAVE: APRENDIZAJE NO SUPERVISADO, DEFUNCIONES FETALES, FACTORES SOCIOECONÓMICOS, COLOMBIA, REDES NEURONALES, MÁQUINA DE BOLTZMANN RESTRINGIDA.

DESCRIPCIÓN:

Las defunciones fetales representan un problema de salud pública en Colombia, asociado a desigualdades sociales y limitaciones en el acceso a servicios de salud. Según el DANE, entre 2020 y 2023 se registraron más de 80.000 casos, lo que evidencia la magnitud del fenómeno. Este proyecto busca identificar las condiciones socioeconómicas que influyen en dichas muertes mediante técnicas de aprendizaje no supervisado. Para ello, se utilizaron datos del DANE y se analizaron variables como edad materna, nivel educativo, afiliación a seguridad social y departamento de residencia. Se aplicó una Máquina de Boltzmann Restringida (RBM) junto con el algoritmo de K-medias, antes y después de la red neuronal, evaluando la homogeneidad de los grupos mediante entropía, con el objetivo de aportar evidencia para fortalecer la vigilancia epidemiológica y orientar políticas públicas en salud materna.

* Trabajo de grado

** Facultad de Ciencias. Escuela de Matemáticas. Director: Andrés Sebastián Ríos Gutiérrez, Candidato a Doctor.

ABSTRACT

TITLE: UNSUPERVISED LEARNING TO IDENTIFY SOCIOECONOMIC FACTORS ASSOCIATED WITH FETAL DEATHS IN COLOMBIA *

AUTHOR: LICETH NATALI BAUTISTA RAMOS - JACKSYMAR PAOLA NUÑEZ ROSALES **

KEYWORDS: UNSUPERVISED LEARNING, FETAL DEATHS, SOCIOECONOMIC FACTORS, COLOMBIA, NEURAL NETWORKS, RESTRICTED BOLTZMANN MACHINE.

DESCRIPTION:

Fetal deaths represent a public health issue in Colombia, associated with social inequalities and limited access to healthcare services. According to DANE, more than 80,000 cases were reported between 2020 and 2023, highlighting the magnitude of the phenomenon. This project aims to identify the socioeconomic conditions influencing these deaths through unsupervised learning techniques. For this purpose, DANE data were used and variables such as maternal age, educational level, social security affiliation, and department of residence were analyzed. A Restricted Boltzmann Machine (RBM) was applied together with the K-means algorithm, both before and after the neural network, evaluating group homogeneity through entropy. The objective is to provide evidence to strengthen epidemiological surveillance and guide public health policies in maternal care.

* Bachelor Thesis

** Faculty of Sciences. School of Mathematics. Advisor: Andrés Sebastián Ríos Gutiérrez, PhD Candidate.

INTRODUCCIÓN

El objetivo de este proyecto es identificar por medio de redes neuronales de clasificación no supervisadas, las condiciones socioeconómicas que pueden influir en las defunciones fetales. A diferencia del aprendizaje supervisado, donde ya se sabe a qué grupo pertenece cada dato, el aprendizaje no supervisado busca encontrar patrones o grupos por sí solo, sin saber de antemano cómo están clasificados los datos. Esta información podría ser útil para guiar futuras intervenciones dirigidas a reducir la tasa de defunciones fetales, por ejemplo, a través de estrategias de educación en salud reproductiva, mejoras en el acceso a la atención médica prenatal y posnatal, o apoyos económicos a poblaciones en riesgo.

Adicionalmente, el análisis se apoya en la construcción de vectores binarios que representan características socioeconómicas de mujeres embarazadas, medidas a partir de cuatro variables seleccionadas para este proyecto: nivel educativo de la madre, afiliación al sistema de seguridad social, edad materna y departamento de residencia. Para cada categoría dentro de estas variables, se creó una columna binaria que toma el valor 1 si la condición está presente, y 0 si no lo está. Por ejemplo, una mujer de 20 a 24 años, con nivel educativo de básica secundaria, afiliada al régimen subsidiado y residente en el departamento del Chocó tendrá un vector con valor 1 en las posiciones correspondientes a esas cuatro categorías, y 0 en las demás. Para la realización de este proyecto se utilizan datos recolectados por el Departamento Administrativo Nacional de Estadística (DANE), en particular los registros de muertes fetales correspondientes a los años 2020, 2021 y 2023. Esta base de datos, de carácter público, ofrece información detallada sobre cada caso de muerte fetal reportado, incluyendo variables como el sitio de ocurrencia, el nivel educativo de la madre, su edad, el tipo de afiliación al sistema de salud y otras variables sociodemográficas relevantes.

Como parte del enfoque metodológico, se utilizó una Máquina de Boltzmann Restringida con el objetivo de capturar las relaciones entre las variables socioeconómicas y la ocurrencia de muerte fetal. Esta red neuronal no supervisada tiene la capacidad de identificar patrones en los datos sin requerir etiquetas previamente definidas. En este estudio, la Máquina de Boltzmann Restringida aprende a representar la información de cada mujer embarazada según sus condiciones socioeconómicas, usando combinaciones de datos codificados en vectores binarios. Así, el modelo puede identificar qué características se repiten con más frecuencia en los casos de muerte fetal. El proceso de entrenamiento de la red se guía por una función de costo, la cual ajusta los pesos para incrementar la probabilidad de representar con precisión los perfiles observados. Los pesos en las conexiones representarían la importancia de cada factor socioeconómico en

el riesgo de muerte fetal, estos problemas de búsqueda podrían representar, por ejemplo, la búsqueda de la mejor manera de mejorar la educación o aumentar el acceso a la atención médica prenatal.

Posteriormente, para interpretar los pesos de las conexiones aprendidos por la Máquina de Boltzmann Restringida, se utilizó un algoritmo de análisis y clasificación llamado K-medias, el cual permitió identificar qué variables tenían mayor importancia en la predicción de defunciones fetales, agrupar observaciones similares y visualizar en las gráficas las respectivas agrupaciones. Además, se aplicó K-medias antes de la Máquina de Boltzmann Restringida con el propósito de contar con un punto de referencia que permitiera comparar los resultados y verificar si la representación generada por la red neuronal mejoraba la agrupación de los datos. Esta aplicación antes y después permitió una mejor visualización y comprensión de los resultados en relación con las variables socioeconómicas.

Finalmente, se calculó la entropía de cada clúster para medir cuán homogéneos eran los grupos internamente en relación con cada variable socioeconómica. Este valor sirvió para analizar qué tan coherentes eran las agrupaciones según la distribución de categorías dentro de cada grupo. Una entropía baja indicaba que los clústeres eran más uniformes, lo que facilitaba su interpretación y descripción, mientras que una entropía alta reflejaba caos, mayor variedad interna y posibles cruces entre grupos.

1. PLANTEAMIENTO

1.1. Descripción del problema

La problemática principal de este trabajo es identificar y analizar los factores socioeconómicos que influyen en las muertes fetales en Colombia. Según el DANE, en el año 2020 se registraron más de 33.000 defunciones¹, mientras que en 2021 fueron alrededor de 30.709², y en 2023 se reportaron 24.862³. Una parte del incremento registrado en 2020 podría estar relacionada con los efectos de la pandemia de COVID-19, la cual impactó negativamente la cobertura y la calidad de la atención en salud materna, limitó el acceso a controles prenatales y profundizó las desigualdades sociales ya existentes. Según el Instituto Nacional de Salud, ese año se observaron retrocesos en varios indicadores de salud perinatal en distintas regiones del país⁴. Esto representa una cifra alarmante, lo que resalta la necesidad de encontrar posibles soluciones que aborden este fenómeno.

1.2. Justificación

Al identificar los factores que inciden en la defunción fetal, como por ejemplo el departamento de residencia de la madre, su nivel educativo, edad y afiliación a la seguridad social, se contribuye al diseño de estrategias que aborden las desigualdades estructurales. Comprender estas relaciones es fundamental para promover intervenciones que impacten directamente en la calidad de vida y el bienestar de las comunidades más vulnerables. En este análisis se eligió usar una Máquina de Boltzmann Restringida porque es un modelo que no necesita datos clasificados previamente y puede aprender patrones a partir de variables categóricas

¹ DANE: *Estadísticas vitales: Defunciones fetales 2020*. En: *Microdatos DANE* (2020). <https://microdatos.dane.gov.co/index.php/catalog/775>.

² DANE: *Estadísticas vitales: Defunciones fetales 2021*. En: *Microdatos DANE* (2021). <https://microdatos.dane.gov.co/index.php/catalog/812>.

³ DANE: *Estadísticas vitales: Defunciones fetales 2023*. En: *Microdatos DANE* (2023). <https://microdatos.dane.gov.co/index.php/catalog/843>.

⁴ Instituto Nacional de Salud (INS): *Informe de Evento: Mortalidad Perinatal y Neonatal Tardía 2020*. En: *INS* (2021). https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA_2020.pdf.

codificadas. Esto resulta útil cuando se quiere encontrar relaciones escondidas en los datos. A diferencia de otros métodos más comunes, la Máquina de Boltzmann Restringida puede captar conexiones complejas entre las variables y simplificar la información sin perder lo más importante, lo que ayuda a identificar los factores que podrían estar más relacionados con la muerte fetal.

1.3. Antecedentes

El uso de técnicas de aprendizaje automático ha cobrado relevancia en el ámbito biomédico en los últimos años, en particular para ayudar a detectar riesgos durante el embarazo. Eugenia Arrieta⁵, desarrolló un sistema de predicción temprana de morbilidad materna extrema utilizando algoritmos supervisados, como regresión logística y máquinas de soporte vectorial. Estos lograron buenos resultados, especialmente en lo que se conoce como alta sensibilidad, que significa que el modelo fue capaz de identificar correctamente a la mayoría de las pacientes que realmente tuvieron complicaciones; este, con datos de la Clínica de Maternidad Rafael Calvo en Cartagena, Colombia. Asimismo, Álvarez, Buitrago, Roa, Tapia y Martínez⁶ diseñaron un aplicativo para identificar el riesgo biopsicosocial en madres gestantes mediante modelos automatizados de aprendizaje, como AutoML, destacando la influencia de factores como el tipo de embarazo y la región de residencia. Este aplicativo consiste en una interfaz que permite al personal médico ingresar los datos clínicos y sociales de la paciente, y recibir como resultado una alerta sobre si presenta o no un riesgo biopsicosocial. El modelo predictivo logró una precisión del 80,7 %, y se entrenó con bases de datos de natalidad y defunciones fetales del DANE entre 2007 y 2016.

No obstante, ambos métodos requieren contar con una variable objetivo establecida, lo que restringe su uso en contextos complejos como la mortalidad fetal, donde las relaciones entre variables pueden ser poco claras y no estar correlacionadas. Por esta razón, este estudio plantea el uso de técnicas de aprendizaje no supervisado, específicamente la Máquina de Boltzmann Restringida, la cual permite descubrir patrones ocultos a partir de variables categóricas sin requerir etiquetas previas. Para comprobar qué tan útil era esta forma de representar los datos, se aplicó el algoritmo K-medias en dos momentos: primero, sobre los datos originales, y luego sobre los datos transformados por la Máquina de Boltzmann Restringida.

⁵ Eugenia Luz Arrieta Rodríguez: *Predicción Temprana de Morbilidad Materna Extrema Usando Aprendizaje Automático*. En: *Universidad Tecnológica de Bolívar* (2017). Tesis de Maestría.

⁶ María Alejandra Álvarez et al.: *Diseño de un aplicativo para la identificación del estado de riesgo biopsicosocial en madres gestantes*. En: *Pontificia Universidad Javeriana* (2019). Trabajo de grado en modalidad de aplicación, Ingeniería Industrial, Primer Semestre 2019.

Esta comparación permitió evaluar si la Máquina de Boltzmann Restringida ayudaba a organizar mejor la información y a formar grupos más claros.

2. OBJETIVOS

2.1. Objetivo general

Identificar las condiciones socioeconómicas que pueden influir en las defunciones fetales en Colombia a través de redes neuronales de clasificación no supervisadas.

2.1.1. Objetivos específicos

1. Describir la base de datos por medio de técnicas descriptivas.
2. Implementar la Máquina de Boltzmann Restringida en la base de datos de defunciones fetales.
3. Aplicar una técnica de análisis de datos como clustering antes y después de entrenar la red con la Máquina de Boltzmann Restringida.
4. Establecer los factores socioeconómicos que influyen en mayor medida sobre las defunciones fetales en Colombia, a través de los pronósticos generados por la técnica aplicada antes y después de la Máquina de Boltzmann Restringida.

3. FUNDAMENTOS DE REDES NEURONALES Y APRENDIZAJE NO SUPERVISADO

3.1. Conceptos básicos

La metodología se apoya en una Máquina de Boltzmann Restringida, que es un tipo de red neuronal artificial. Para entender mejor este concepto, primero es necesario definir qué es una red neuronal de forma general.

3.2. Redes neuronales

⁷ Una red neuronal artificial, se define como un modelo computacional diseñado para procesar información de manera paralela y distribuida. Su estructura busca imitar las funciones básicas del sistema nervioso humano mediante la conexión de múltiples unidades de procesamiento (denominadas nodos o neuronas artificiales) que operan de forma local y colaboran a través de un mecanismo de aprendizaje conjunto. El proceso mediante el cual estas neuronas ajustan su comportamiento para cumplir un objetivo específico se conoce como algoritmo de aprendizaje. Este se encarga de actualizar de manera sistemática los pesos sinápticos de la red para optimizar su rendimiento en la tarea asignada.

⁸ Las redes neuronales artificiales (RNA) surgieron como un intento de emular el funcionamiento de las neuronas del cerebro humano. En este sentido, las RNA siguen una línea distinta a los enfoques clásicos de la inteligencia artificial, los cuales buscan modelar la inteligencia humana imitando los procesos de razonamiento que ocurren en el cerebro. En esta introducción, se presentan los conceptos iniciales y fundamentales relacionados con las RNA.

⁹ Como en el caso de la neurona biológica, la neurona artificial recibe unas entradas de estímulo que

⁷ Universidad de Sevilla: *Capítulo 4. En Proyecto final de carrera*. En: *Recurso en línea* (s.f.). <https://biblus.us.es/bibing/proyectos/use/abreproy/10807/fichero/Cap%C3%ADtulo+4.pdf>.

⁸ E. F. Caicedo y J. A. López: *Una aproximación práctica a las redes neuronales artificiales*. En: *Programa Editorial Universidad del Valle* (2017).

⁹ E. F. Caicedo y J. A. López: *Una aproximación práctica a las redes neuronales artificiales*. En: *Programa*

pueden provenir del sistema sensorial externo o de otras neuronas con las cuales posee conexión. La información que recibe la neurona la definimos con el vector de entradas $\mathbf{X} = [X_1, X_2, \dots, X_n]$. La información recibida por la neurona es modificada por un vector \mathbf{B} de pesos sinápticos cuyo papel es el de emular la sinapsis existente entre las neuronas biológicas. Estos pesos pueden interpretarse como factores que regulan la importancia de cada entrada: si un peso es alto, la entrada correspondiente tendrá mayor influencia en la activación de la neurona; si es bajo o negativo, su efecto será reducido o inverso. Además, se añade un parámetro denominado sesgo o umbral, representado por b , el cual permite ajustar la salida de la neurona incluso cuando las entradas son nulas. A partir de la combinación de las entradas ponderadas por sus pesos y del sesgo, se calcula lo que se conoce como la entrada neta. Esta entrada neta es posteriormente evaluada por una función de activación, que determina si la neurona se activa y con qué intensidad.

¹⁰ La máquina de Boltzmann es un modelo de red neuronal que se utiliza para clasificar datos. En el contexto de las defunciones fetales, una red neuronal de clasificación podría utilizarse para predecir si una mujer embarazada tiene un mayor riesgo de muerte fetal.

Definición 3.2.1. ¹¹ **Red monocapa:** *Una red neuronal monocapa es un tipo de red artificial compuesta por una sola capa de neuronas entre la entrada y la salida, definida por $\mathbf{y} = f(\mathbf{B}\mathbf{X} + \mathbf{b})$. Consta de dos niveles de neuronas, el nivel de entrada que tiene los datos de entrada de la red y el nivel de salida que tiene la salida de la red, el primero asociado al vector de entrada, pero no hay ningún tipo de procesamiento de estos datos, por esta razón no se considera formalmente como una capa, únicamente sirve de interfaz entre los datos de entrada y las siguientes capas o neuronas.*

Editorial Universidad del Valle (2017).

¹⁰ E. F. Caicedo y J. A. López: *Una aproximación práctica a las redes neuronales artificiales*. En: *Programa Editorial Universidad del Valle (2017)*.

¹¹ E. F. Caicedo y J. A. López: *Una aproximación práctica a las redes neuronales artificiales*. En: *Programa Editorial Universidad del Valle (2017)*.

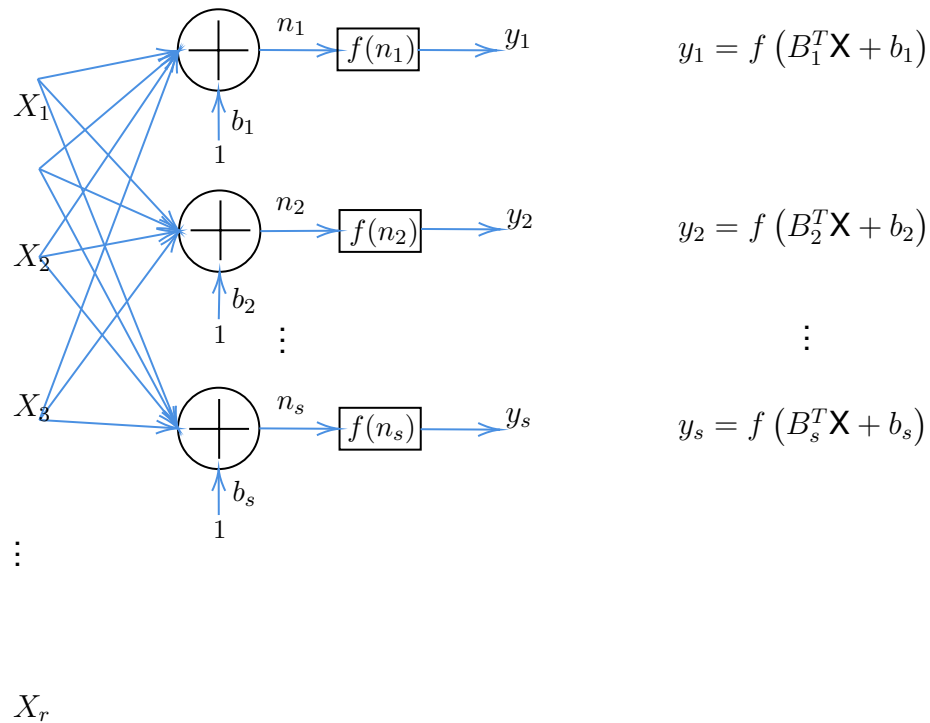


Figura 1. Red neuronal Monocapa.

Dónde,

- La \mathbf{X} representa el vector de entradas, con una dimensión $r \times 1$.
- La \mathbf{b} representa el vector del sesgo, con s sesgos hay s entradas.
- La \mathbf{n} representa el vector de entrada neta, es decir, los valores resultantes de combinar linealmente las entradas ponderadas por sus pesos y sumadas con el sesgo. Cada componente n_s se calcula como: $n_s = \mathbf{B}_s^T \mathbf{X} + b_s$ y determina lo que recibe cada neurona antes de aplicar la función de activación. Tiene dimensión $s \times 1$
- La \mathbf{y} representa el vector de salida, con una dimensión $s \times 1$.
- La matriz de pesos sinápticos denotada con \mathbf{B} contiene los vectores de los pesos sinápticos de cada neurona, dónde la fila 1 corresponde a los pesos sinápticos de la neurona 1 y así sucesivamente hasta llegar a la neurona s . La primer columna representa los pesos sinápticos conectados a la

entrada 1, y así sucesivamente. Cualquier peso sináptico puede estar representado por un B_{ij} donde i es el número de neurona y j el número de entrada.

$$\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_r \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{bmatrix} \quad \mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_s \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_s \end{bmatrix} \quad (1)$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1^T \\ \mathbf{B}_2^T \\ \vdots \\ \mathbf{B}_S^T \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1R} \\ B_{21} & B_{22} & \cdots & B_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ B_{S1} & B_{S2} & \cdots & B_{SR} \end{bmatrix} \quad (2)$$

La siguiente ecuación matricial y vectorial representa a la red neuronal monocapa

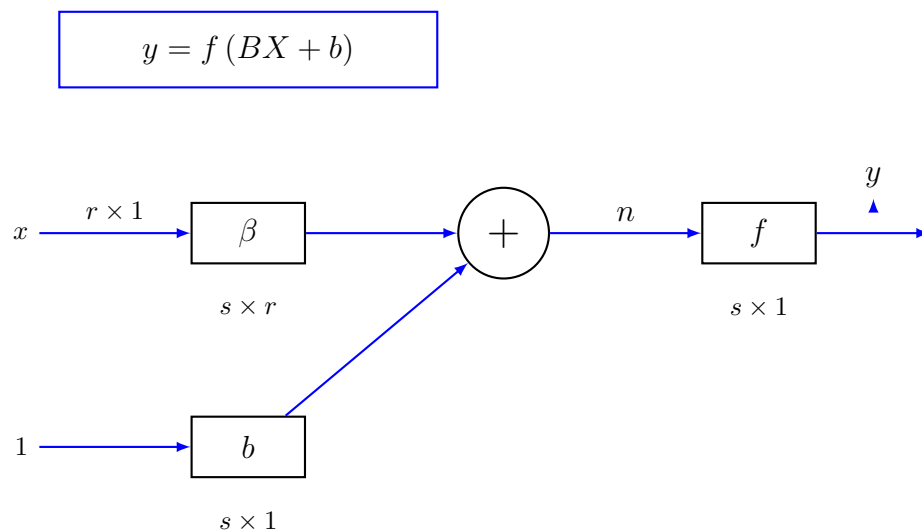


Figura 2. Representación matricial y vectorial de la red neuronal monocapa.

Definición 3.2.2. ¹² **Red multicapa:** Una red neuronal multicapa se define mediante:

$$\begin{aligned} \mathbf{y}^{(1)} &= f^{(1)} \left(\mathbf{B}^{(1)} \mathbf{X} + \mathbf{b}^{(1)} \right), \\ \mathbf{y}^{(k)} &= f^{(2)} \left(\mathbf{B}^{(2)} \mathbf{y}^{(1)} + \mathbf{b}^{(2)} \right), \\ \hat{\mathbf{y}} &= f^{(3)} \left(\mathbf{B}^{(3)} \mathbf{y}^{(2)} + \mathbf{b}^{(3)} \right), \end{aligned}$$

donde cada $\mathbf{B}^{(l)}$ representa la matriz de pesos sinápticos correspondiente a la capa l , $\mathbf{b}^{(l)}$ es el vector de sesgos de dicha capa, y $f^{(l)}$ la función de activación asociada.

La entrada neta de la capa l se denota por:

$$\mathbf{n}^{(l)} = \mathbf{B}^{(l)} \mathbf{y}^{(l-1)} + \mathbf{b}^{(l)},$$

y la salida correspondiente es:

$$\mathbf{y}^{(l)} = f^{(l)}(\mathbf{n}^{(l)}).$$

Aquí, $\mathbf{y}^{(1)}$ representa la salida de la primera capa oculta, $\mathbf{y}^{(k)}$ la de una capa oculta intermedia, y $\hat{\mathbf{y}}$ es la salida final de la red.

En este caso, la red posee un nivel de entrada con X neuronas y una capa de salida con y neuronas. A diferencia de la red monocapa, se incorpora al menos una capa intermedia entre la entrada y la salida, denominada capa oculta, compuesta por h neuronas.

¹³ Las capas ocultas se encargan del procesamiento de los datos que reciben de otras neuronas, aplicando sobre ellos una cierta transformación $f(x)$, generalmente no-lineal. El número de capas ocultas es variable y depende del método en particular. Se considera como mínimo una capa oculta, sin existir límite máximo.

¹² E. F. Caicedo y J. A. López: *Una aproximación práctica a las redes neuronales artificiales*. En: Programa Editorial Universidad del Valle (2017).

¹³ J. Cantero Lorenzo: *Máquina de aprendizaje y aplicaciones*. En: Universitat de les Illes Balears (2022).

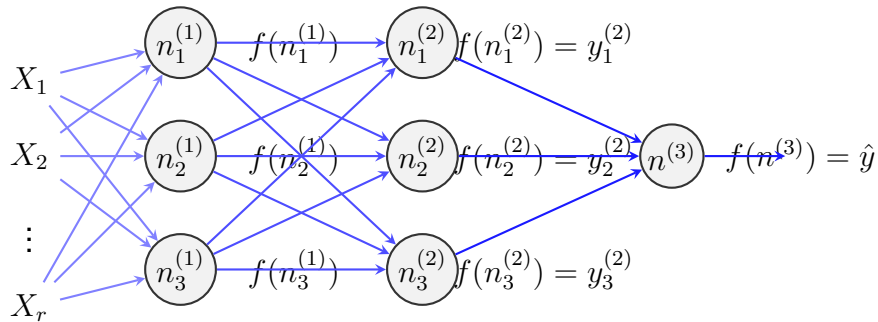


Figura 3. Red neuronal multicapa.

Notación	Significado
X_r	Entrada r -ésima a la red
$n_j^{(1)}$	Entrada neta del nodo j en la primera capa oculta
$f(n_j^{(1)})$	Salida activada del nodo j en la primera capa oculta
$n_k^{(2)}$	Entrada neta del nodo k en la segunda capa oculta
$y_k^{(2)} = f(n_k^{(2)})$	Salida activada de la neurona k en la segunda capa oculta
$n^{(3)}$	Entrada neta de la neurona en la capa de salida
$y = f(n^{(3)})$	Salida final de la red neuronal multicapa
$\mathbf{B}^{(l)}$	Matriz de pesos sinápticos correspondiente a la capa l
$\mathbf{b}^{(l)}$	Vector de sesgos correspondiente a la capa l

Tabla 1. Notación empleada en la red neuronal.

¹⁴ **Función de activación:** Una función de activación es un elemento matemático que se encarga de

¹⁴ Simon Haykin: *Neural Networks and Learning Machines*. 3rd. Pearson Education, 2009.

transformar la entrada neta de una neurona para generar su salida. Su función principal consiste en incorporar no linealidad al modelo, lo que permite a la red capturar patrones complejos en los datos. Si no se utilizaran funciones de activación no lineales, incluso una red profunda actuaría como un modelo lineal, limitando su capacidad para representar relaciones complejas.¹⁵ Las funciones de activación desempeñan un papel esencial en las redes neuronales artificiales, ya que transforman las combinaciones lineales de las entradas y pesos en señales de salida que se transmiten a las siguientes capas. Estas funciones permiten que la red aprenda patrones complejos y realice predicciones más precisas. Aunque no hay reglas fijas sobre la cantidad óptima de capas o la función de activación ideal, se reconoce que utilizar múltiples capas ocultas y seleccionar adecuadamente las funciones de activación puede mejorar significativamente el rendimiento de la red.

¹⁶ Cuando tenemos una función de activación, es crucial considerar el clasificador basado en umbral. Esto implica decidir si el valor de la transformación lineal activa o no la neurona es decir, una neurona se activa si la entrada a la función de activación es mayor que un valor de umbral, de lo contrario, se desactiva. En ese caso, la salida no se alimenta como entrada a la siguiente capa.

- **Función escalon:**¹⁷ La Función Escalón o Función Escalón Binaria es la forma más básica de función de activación, puede implementarse fácilmente con simples declaraciones if-else en Python. Cuando creamos un clasificador binario, es común utilizar funciones de activación binarias. Sin embargo, la función de escalón binario no es adecuada para la clasificación multiclase en objetivos múltiples. Además, el gradiente de esta función es cero, lo que puede obstaculizar el proceso de Backpropagation ¹⁸. Es decir, al calcular la derivada de $f(x)$ con respecto a x , obtendremos un valor

¹⁵ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

¹⁶ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

¹⁷ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

¹⁸ El algoritmo de backpropagation es una técnica utilizada para entrenar redes neuronales multicapa. Consiste en propagar la información hacia adelante para obtener una salida, y luego ajustar los pesos y sesgos hacia atrás minimizando el error mediante descenso por gradiente.

nulo.

$$F(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1, & \text{si } x \geq 0 \end{cases}$$

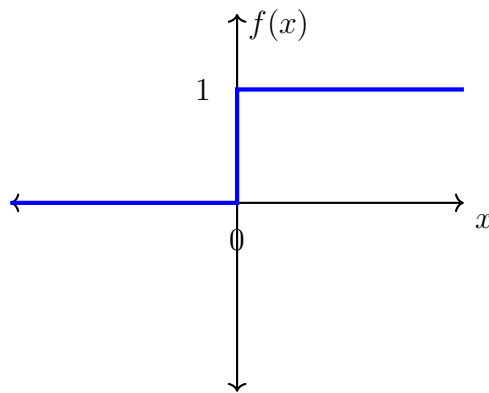


Figura 4. Función escalón

- **Función de activación lineal:**¹⁹ La función de activación lineal es proporcional a su entrada y se define como $F(x) = ax$, donde a es una constante elegida por el usuario. A diferencia de la función de paso binario, cuyo gradiente es cero, la función lineal tiene un gradiente constante, lo que permite actualizar pesos y sesgos durante el *Backpropagation*. Sin embargo, esta función presenta limitaciones, ya que no mejora la precisión de la red ni captura patrones complejos, siendo más adecuada para tareas simples que priorizan la interpretabilidad.

¹⁹ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

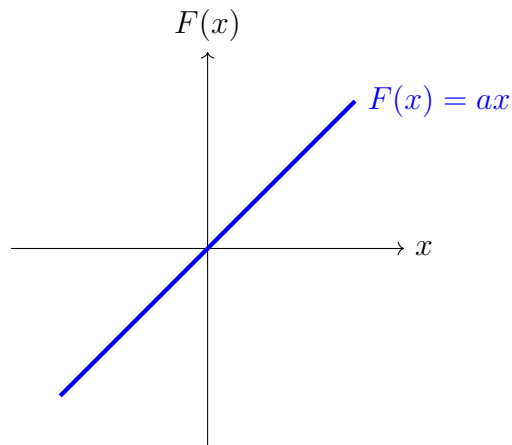


Figura 5. Función de activación lineal

- Función de activación sigmoide:** La función sigmoide, definida como $f(x) = \frac{1}{1+e^{-x}}$, es ampliamente utilizada como función de activación debido a su naturaleza no lineal, suavidad y forma de S continua, siendo continuamente diferenciable. Su derivada es $f'(x) = (1 - (1 + e^{-x}))$. Aunque no es simétrica respecto a cero, lo que hace que los signos de las salidas de las neuronas sean iguales, este inconveniente puede mitigarse escalándola adecuadamente.²⁰

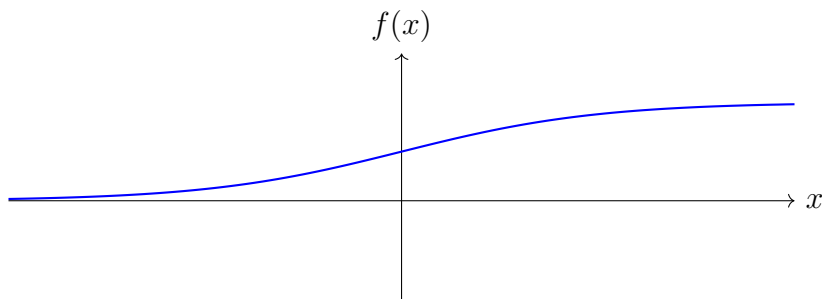


Figura 6. Función sigmoide

- Función Tangente Hipérbolica:** La función Tangente Hiperbólica, definida como $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, es similar a la función sigmoide, pero simétrica respecto al origen, lo que permite que las salidas tengan signos diferentes como entrada en la siguiente capa. Es continua, diferenciable y su rango está entre

²⁰ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

-1 y 1. En comparación con la sigmoide, tiene un gradiente más pronunciado, lo que la hace preferible al no restringir la variación del gradiente a una sola dirección y estar centrada en cero.²¹

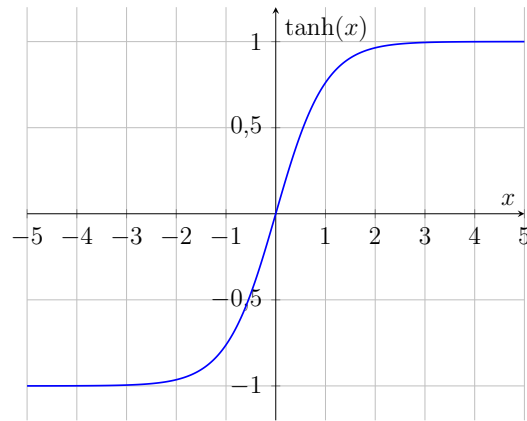


Figura 7. Función Tangente Hiperbólica

- **Función de activación ReLU:**²² La función ReLU (*unidad lineal rectificada*), definida como $f(x) = \max(0, x)$, es una función de activación no lineal ampliamente utilizada en redes neuronales por su eficiencia y capacidad para capturar patrones complejos. Su característica principal es que desactiva neuronas cuando la entrada es negativa, lo que genera activaciones dispersas y computación eficiente. Aunque tiene un punto no diferenciable en $x = 0$, su no linealidad permite a las redes aprender representaciones complejas.

²¹ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

²² Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

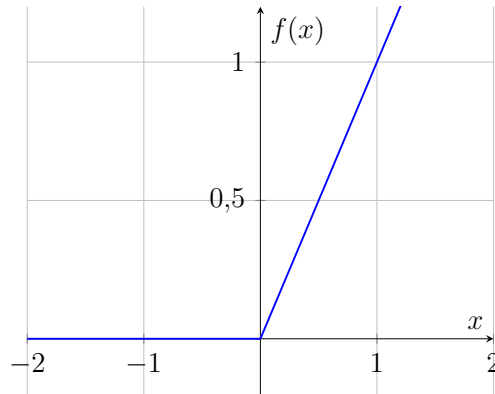


Figura 8. Función ReLU

- Función Leaky ReLU:**²³ La función Leaky ReLU es una variante de ReLU que introduce una pequeña pendiente en la región negativa, evitando que las entradas negativas se anulen completamente. Se define como $f(x) = x$ si $x \geq 0$ y $f(x) = 0,01 \times x$ si $x < 0$, donde 0,01 es un valor positivo que controla la pendiente en valores negativos. Esta modificación mejora el flujo del gradiente y ayuda a mitigar el problema de las neuronas muertas.

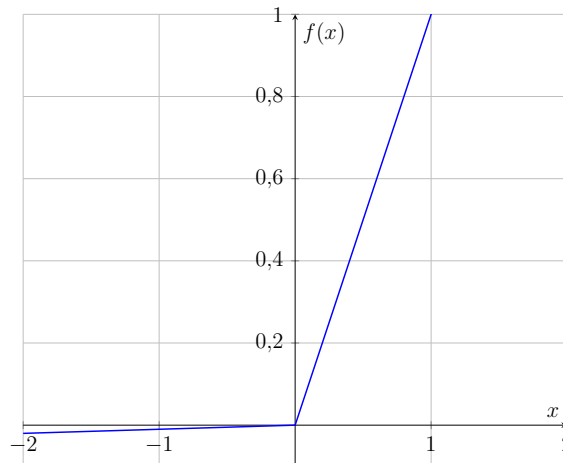


Figura 9. Función Leaky ReLU

²³ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

- **Función ReLU parametrizada:**²⁴ También es una variante de la Unidad Lineal Rectificada (ReLU) con un rendimiento mejorado y una pequeña variación. Resuelve el problema del gradiente de ReLU que se vuelve cero para valores negativos de X al introducir un nuevo parámetro para la parte negativa de la función, es decir, la pendiente. Cuando el valor de a se establece en 0.01, la función se comporta como la función Leaky ReLU, pero aquí a también es un parámetro entrenable. Para lograr una convergencia más rápida y óptima, la red neuronal aprende el valor de a . Se expresa como:

$$f(x) = \begin{cases} x, & \text{si } x \geq 0 \\ ax, & \text{si } x < 0 \end{cases}$$

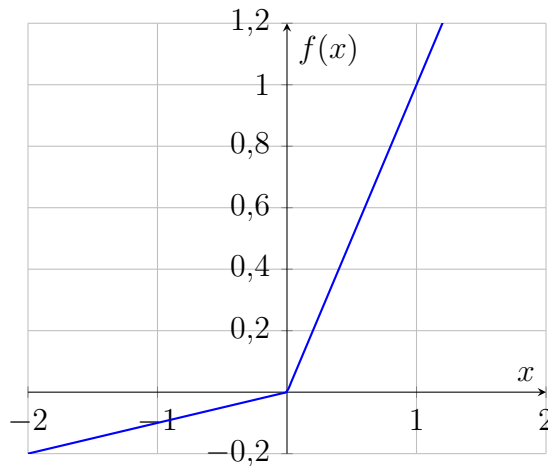


Figura 10. Función ReLU

- **Función Unidad Exponencial Lineal:**²⁵ La Unidad Exponencial Lineal (ELU) es una variante de ReLU que utiliza una curva logarítmica para valores negativos.²⁶ A diferencia de las ReLU, las ELU permiten valores negativos, alineando gradientes y reduciendo el sesgo al llevar activaciones hacia

²⁴ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

²⁵ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

²⁶ Djork-Arne Clevert, Thomas Unterthiner y Sepp Hochreiter: *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. En: *International Conference on Learning Representations (ICLR)* (2016).

cero, similar a la normalización por lotes pero con menor carga computacional. Además, las ELU se saturan a valores negativos con entradas pequeñas, disminuyendo la variación y el ruido en la transmisión de información. Se define como:

$$f(x) = \begin{cases} x, & \text{si } x \geq 0 \\ a(e^x - 1), & \text{si } x < 0 \end{cases}$$

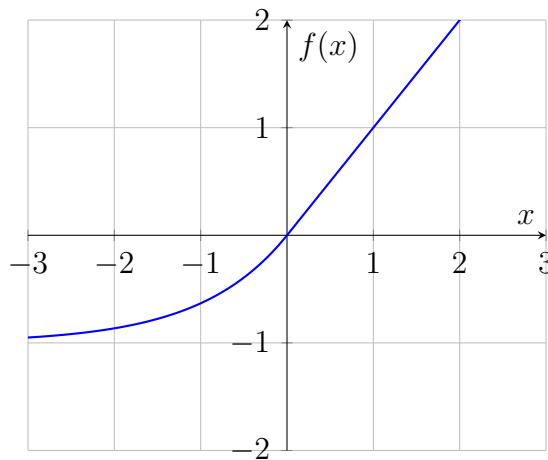


Figura 11. Función ELU

- **Función Swish:**²⁷ La función Swish es una función de activación relativamente nueva que fue descubierta por investigadores. La característica distintiva de la función Swish es que no es monótona, lo que significa que el valor de la función puede disminuir incluso cuando los valores de las entradas están aumentando. En algunos casos, Swish supera incluso a la función ReLU. Se expresa matemáticamente como:

$$f(x) = \frac{x}{1 + e^{-x}}$$

- **Función SoftMax:**²⁸ La función *Softmax* es una generalización de la función sigmoide diseñada para

²⁷ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

²⁸ Siddharth Sharma, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

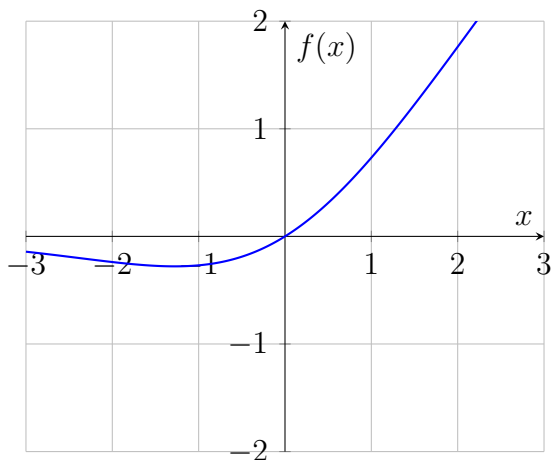


Figura 12. Función Swish

clasificación multiclase. Mientras que la sigmoide se emplea comúnmente en problemas binarios, la Softmax permite obtener una distribución de probabilidad sobre varias clases, asignando a cada una una probabilidad entre 0 y 1, cuya suma total es 1. Esta función transforma los valores de salida de una red neuronal en probabilidades relativas a cada clase, y se define como:

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$$

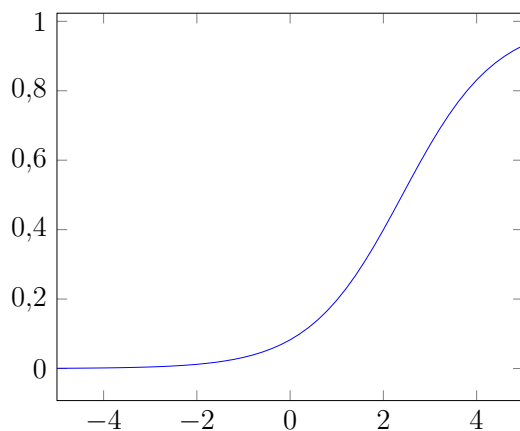


Figura 13. Función Softmax

Las redes neuronales, un tipo de modelo de Machine Learning ampliamente utilizado, emplean dos

enfoques principales para el aprendizaje: el aprendizaje supervisado y el aprendizaje no supervisado.

Aprendizaje supervisado:²⁹ Los algoritmos de aprendizaje supervisado están basados en un modelo predictivo, el cual está compuesto por dos grupos de datos uno para realizar el entrenamiento, otro de prueba y un mecanismo que permita evaluar si el algoritmo está haciendo las cosas bien. El conjunto de datos previamente etiquetado y clasificado, del cual ya se sabe a qué grupo, valor o categoría pertenecen son los datos de entrenamiento y son utilizados para realizar el ajuste al modelo, los datos son utilizados por el algoritmo para ir “aprendiendo” a clasificar las muestras realizando una comparación del resultado obtenido por el modelo, y el valor inicial de la muestra, realizando compensaciones con respecto al modelo y a cada error en la estimación del resultado.

Aprendizaje no supervisado:³⁰ En esta tesis se utilizará el enfoque de aprendizaje no supervisado, el cual se diferencia del modelo anterior porque solo tiene en cuenta los datos de entrada para realizar sus ajustes al modelo predictivo, sin importar si están o no clasificados o etiquetados, ya que estas características no son necesarias para realizar el entrenamiento del modelo. En el aprendizaje no supervisado se puede ver que las etiquetas de salida no son entregadas a los algoritmos, sino que tienen que ser identificadas.

³¹Las redes de aprendizaje no supervisado no requieren influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta a una determinada entrada es o no correcta. Estas redes están diseñadas para descubrir patrones, regularidades, relaciones o categorías presentes en los datos de entrada. La interpretación de la salida de estas redes puede variar según su estructura y el algoritmo de aprendizaje utilizado. En ocasiones, la salida indica el nivel de familiaridad o similitud entre la información presentada en la entrada y los datos previamente vistos.

²⁹ M. German Ruiz: *Modelo de análisis de datos utilizando técnicas de aprendizaje supervisado y no supervisado, para identificar patrones en la información generada por los pacientes, sometidos a juegos diseñados como un instrumento de apoyo terapéutico*. En: *Universidad Jorge Tadeo Lozano, Facultad de Ciencias Naturales e Ingeniería* (2019).

³⁰ M. German Ruiz: *Modelo de análisis de datos utilizando técnicas de aprendizaje supervisado y no supervisado, para identificar patrones en la información generada por los pacientes, sometidos a juegos diseñados como un instrumento de apoyo terapéutico*. En: *Universidad Jorge Tadeo Lozano, Facultad de Ciencias Naturales e Ingeniería* (2019).

³¹ Damián Jorge Matich: *Redes Neuronales: Conceptos Básicos y Aplicaciones*. En: *Universidad Tecnológica Nacional – Facultad Regional Rosario* (2001).

3.3. Redes neuronales

El perceptrón simple es un modelo de red neuronal artificial que fue introducido por Frank Rosenblatt en 1962. Es un modelo unidireccional, lo que significa que los datos fluyen en una sola dirección, de la capa de entrada a la capa de salida. Un perceptrón simple consta de dos capas de neuronas: una capa de entrada y una capa de salida. El perceptrón consta de una capa de entrada con r características y una capa de salida con s neuronas. Este modelo obtiene al aplicar la función de activación a la combinación lineal de las entradas ponderadas y el sesgo correspondiente. Este modelo constituye la base de redes neuronales más complejas, como las redes multicapa.

Definición 3.3.1.³² **Perceptrón simple:** Sea la entrada correspondiente al vector $\mathbf{X} = [X_{1,i}, X_{2,i}, \dots, X_{r,i}]$, donde $X_{d,i}$ es la d -ésima variable de entrada del i -ésimo individuo con $d = 1, \dots, r$, y cuya salida corresponde a y_i para cada $i = 1, \dots, n$. Sea el modelo

$$y_i = f \left(\sum_{j=1}^r B_{ij} X_{j,i} + b_i \right) \quad \text{para } i = 1, \dots, n \quad (3)$$

donde $\mathbf{B} = [B_{ij}]$ se denomina como la matriz de pesos sinápticos de dimensión $s \times r$, $\mathbf{b} = [b_1, b_2, \dots, b_s]$ como el vector de sesgos y $f(\cdot)$ como la función de activación. El modelo dado por la ecuación 3 se define como **red neuronal de perceptrón simple**.

Definición 3.3.2.³³ **Perceptrón multicapa:** Sean $\mathbf{X} = [x_1, x_2, \dots, x_n]$ el vector de entrada; $\mathbf{y}^{(1)}$ el vector de salidas de la capa oculta; y \hat{y} la salida final de la red. Los pesos sinápticos de la capa oculta se representan mediante la matriz $\mathbf{B}^{(1)}$, y su vector de sesgos correspondiente por $\mathbf{b}^{(1)}$. De manera similar, $\mathbf{B}^{(2)}$ representa los pesos de la capa de salida y $\mathbf{b}^{(2)}$ sus sesgos. La operación de un perceptrón multicapa se expresa como:

$$\hat{y}_k = \sum_{j=1}^q B_{kj}^{(2)} \cdot f^{(1)} \left(\sum_{i=1}^n B_{ji}^{(1)} \cdot X_i + b_j^{(1)} \right) + b_k^{(2)}$$

³² Pedro Larrañaga, Iñaki Inza y Abdelmalik Moujahid: *Redes Neuronales*. En: *Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad del País Vasco–Euskal Herriko Unibertsitatea* (s.f.).

³³ Pedro Larrañaga, Iñaki Inza y Abdelmalik Moujahid: *Redes Neuronales*. En: *Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad del País Vasco–Euskal Herriko Unibertsitatea* (s.f.).

Donde $\mathbf{X} \in \mathbb{R}^{n \times 1}$ representa el vector de entradas, $\mathbf{B}^{(1)} \in \mathbb{R}^{q \times n}$ son los pesos sinápticos de la capa oculta, $\mathbf{b}^{(1)} \in \mathbb{R}^{q \times 1}$ es el vector de sesgo de dicha capa, y $f^{(1)}(\cdot)$ es su función de activación. Por otro lado, $\mathbf{B}^{(2)} \in \mathbb{R}^{m \times q}$ corresponde a los pesos sinápticos de la capa de salida, $\mathbf{b}^{(2)} \in \mathbb{R}^{m \times 1}$ es el sesgo asociado, $f^{(2)}(\cdot)$ es su función de activación, y $\hat{\mathbf{y}} \in \mathbb{R}^{m \times 1}$ es la salida final de la red neuronal.

Esta salida puede compararse posteriormente con un vector objetivo \mathbf{t} para evaluar el error y ajustar los parámetros de la red durante el entrenamiento.

3.4. Modelos no supervisados

³⁴ En el aprendizaje no supervisado, la máquina solo recibe entradas x_1, x_2, \dots, x_t , sin recibir salidas supervisadas o recompensas de su entorno. La falta de retroalimentación puede parecer misteriosa en cuanto a lo que la máquina puede aprender. Sin embargo, es posible establecer un marco formal para el aprendizaje no supervisado. Este se basa en la premisa de que la máquina busca construir representaciones de las entradas que sean útiles para la toma de decisiones, la predicción de futuras entradas, la comunicación eficiente con otras máquinas, entre otros propósitos.

En esencia, el aprendizaje no supervisado implica descubrir patrones en los datos que van más allá del ruido no estructurado o variabilidad aleatoria en los datos. Ejemplos clásicos incluyen la agrupación o clustering y la reducción de dimensionalidad.

- **Clustering:**³⁵Un cluster de puntos no es un objeto bien definido. En su lugar, los cluster o agrupaciones se definen en función de su entorno y la escala en la que se examinan los datos. Dos posibles definiciones para cluster son: **I)** Los patrones dentro de un cluster son más similares entre sí que los patrones pertenecientes a otros grupos. **II)** Un clúster es una región de puntos de alta densidad, separada de otros grupos por zonas de baja densidad relativa.

³⁶Este conjunto de algoritmos sigue un enfoque iterativo, comenzando con la premisa de que cada punto de datos constituye un grupo independiente. En cada iteración, dos grupos se combinan hasta

³⁴ O. Bousquet, U. von Luxburg y G. Rätsch: *Advanced Lectures on Machine Learning*. Springer, 2003.

³⁵ O. Bousquet, U. von Luxburg y G. Rätsch: *Advanced Lectures on Machine Learning*. Springer, 2003.

³⁶ O. Bousquet, U. von Luxburg y G. Rätsch: *Advanced Lectures on Machine Learning*. Springer, 2003.

alcanzar un número predeterminado de grupos. La selección de qué grupos fusionar se basa en diversas métricas, como la distancia entre los centros de los grupos o la distancia entre los puntos más cercanos de grupos diferentes. Aunque el Algoritmo de Agrupamiento Jerárquico Aglomerativo (AGHC) es fácil de comprender y aplicar, su alto consumo computacional lo hace poco práctico para conjuntos de datos medianos y grandes.

- **Reducción de dimensionalidad:**³⁷ Parece razonable pensar que recopilar más características descriptivas de los datos debería facilitar su clasificación. Sin embargo, un exceso de datos y variables puede ser contraproducente, especialmente cuando estos están altamente correlacionados con los ya existentes. Las características ruidosas e irrelevantes ³⁸, lejos de ser útiles, resultan perjudiciales para la clasificación. Su presencia aumenta la complejidad computacional, exige un mayor número de muestras de entrenamiento para alcanzar una precisión determinada y, en última instancia, degrada la capacidad de generalización del clasificador, porque, cuando se incluyen variables que no aportan información útil, como las irrelevantes, el modelo se vuelve más complejo y necesita procesar más datos, lo que puede causar que el modelo se ajuste demasiado a los datos conocidos y no funcione bien con datos nuevos, afectando su rendimiento. En esencia, la calidad de los datos, no la cantidad, es el factor determinante para una clasificación exitosa.

El objetivo de la reducción de dimensionalidad es identificar un subconjunto de variables, o combinaciones de estas, que representen los datos de forma eficiente. Este proceso incluye dos enfoques principales: Selección de características, que consiste en conservar solo las variables relevantes y eliminar las redundantes. Esto permite reducir el ruido presente en los datos, es decir, eliminar variaciones aleatorias que no aportan información útil, lo cual mejora la estabilidad del modelo y su capacidad para aprender patrones reales; y Reducción de características, que busca representar las características originales mediante nuevas combinaciones más compactas, sin perder información esencial.

En esta tesis, la reducción de dimensionalidad cumple un papel fundamental como etapa previa a la clasificación no supervisada mediante K-medias. Para ello, se emplea una Máquina de Boltzmann Restringida, la cual extrae representaciones más estructuradas y compactas del conjunto original

³⁷ O. Bousquet, U. von Luxburg y G. Rätsch: *Advanced Lectures on Machine Learning*. Springer, 2003.

³⁸ En la reducción de dimensionalidad, se consideran características ruidosas aquellas que introducen variaciones aleatorias no informativas, y características irrelevantes las que no aportan valor al análisis. Ambas pueden perjudicar el rendimiento del modelo y deben ser filtradas para mejorar su precisión y eficiencia

de datos. Este enfoque permite disminuir la redundancia y filtrar información irrelevante, lo que se traduce en una mejora del rendimiento de la agrupación.

4. ALGORITMOS A IMPLEMENTAR

4.1. Máquina de Boltzmann Restringida

³⁹ La Máquina de Boltzmann es una máquina de estados: repite o modifica ciertos eventos hasta alcanzar un estado considerado estable. Estos eventos consisten en ponderaciones matemáticas que se ajustan a un patrón de valores. Cuando estas ponderaciones se estabilizan dentro de un margen confiable, se considera que la máquina ha aprendido ese patrón matemáticamente. Este fenómeno se puede comparar con lo que ocurre físicamente cuando un material recibe energía (calor), se dilata y aumenta su superficie. La energía del sistema físico puede representarse como:

$$E = - \sum_{i < j} B_{ij} \cdot X_i \cdot X_j + \sum_i b_i \cdot X_i \quad (4)$$

dónde la variación de energía al modificar el estado de una unidad k se expresa como:

$$\Delta E_k = \sum_i B_{ki} \cdot X_i - b_k \quad (5)$$

- B_{ij} : fuerza de conexión (peso sináptico) entre las unidades i y j .
- X_i : estado (entrada) binario de la unidad i .
- b_i : umbral o sesgo asociado a la unidad i .

⁴⁰ Dentro de este umbral, una determinada salida de energía puede aceptarse o no como señal de cambio de estado del sistema completo. En 1982, Hopfield propuso una red neuronal artificial en la que, si las unidades estaban conectadas simétricamente ($B_{ij} = B_{ji}$) y los pesos se actualizaban simultáneamente,

³⁹ Daniel Nehme Müller: *Máquina de Boltzmann*. En: *Universidade Federal do Rio Grande do Sul* (1993). https://danielnm.gitlab.io/inf-ufrgs/docs/t1_arqe.pdf.

⁴⁰ Daniel Nehme Müller: *Máquina de Boltzmann*. En: *Universidade Federal do Rio Grande do Sul* (1993). https://danielnm.gitlab.io/inf-ufrgs/docs/t1_arqe.pdf.

era posible simular la reducción de energía (entropía) de un sistema físico para modelar un patrón.

Las Máquinas de Boltzmann Restringidas son un tipo de red neuronal no supervisada que se diferencia de las redes tradicionales porque no tienen conexiones entre neuronas de la misma capa, lo que simplifica su arquitectura. A diferencia de las redes supervisadas, la máquina de Boltzmann no requiere etiquetas para entrenarse: aprenden a modelar la distribución de probabilidad de los datos mediante un proceso de muestreo estocástico.

4.1.0.1. Modelo gráfico y parametrización⁴¹ La máquina de Boltzmann es una red neuronal artificial estocástica generativa capaz de aprender una distribución de probabilidad sobre sus conjuntos de entradas. Posee un grafo bipartito ⁴² es decir, unidades visibles y ocultas sin conexiones intracapa, por lo que se encuentran restringidas. La restricción es que sus unidades neuronales que conectan unidades visibles y ocultas forman un gráfico bipartito, mientras que no se permite ninguna conexión con el mismo grupo (visible u oculto).

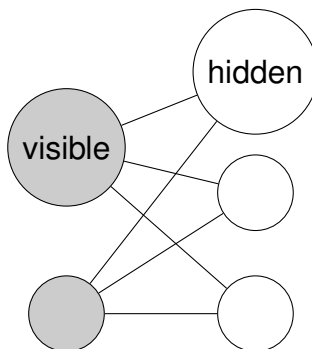


Figura 14. Estructura de una RBM con dos nodos visibles y tres nodos ocultos.

⁴³Los nodos o neuronas son variables aleatorias cuyos estados dependen del estado de los demás nodos

⁴¹ Mohamed A. El-Brawany et al.: *Artificial intelligence-based data-driven prognostics in industry: A survey*. En: *Elsevier* 184 (2023).

⁴² (S. Wasserman y K. Faust: *Social Network Analysis: Methods and Applications*. Cambridge University Press, 2013)Un grafo bipartito es aquel en el que puedes separar todos los nodos en dos grupos diferentes de manera que cada conexión une un nodo de un grupo con uno del otro. No existen conexiones entre nodos del mismo grupo.

⁴³ Geoffrey Hinton: *A Practical Guide to Training Restricted Boltzmann Machines*. En: *Department of*

a los que están conectados. Por lo tanto, el modelo está parametrizado por los pesos de las conexiones, así como un término de intersección (sesgo) para cada unidad visible y oculta.

La función energía mide la calidad de una asignación conjunta:

$$E(\mathbf{X}, \mathbf{n}^{(1)}) = \sum_i b_i X_i - \sum_j b_j^{(1)} n_j^{(1)} - \sum_{i,j} X_i \cdot n_j^{(1)} \cdot B_{ji}^{(1)} \quad (6)$$

Donde:

- X_i : estado binario de la unidad visible i
- $n_j^{(1)}$: estado binario de la unidad oculta j
- b_i : sesgo de la unidad visible i
- $b_j^{(1)}$: sesgo de la unidad oculta j
- $B_{ji}^{(1)}$: peso sináptico entre la unidad visible i y la oculta j

⁴⁴ La red asigna una probabilidad a cada posible par de vectores de entrada y salida a través de esta función de energía:

$$p(\mathbf{x}, \mathbf{n}^{(1)}) = \frac{1}{Z} e^{-E(\mathbf{x}, \mathbf{n}^{(1)})} \quad (7)$$

⁴⁵ donde la “función de partición” Z viene dada por la suma de todos los posibles pares de vectores de entrada y salida:

Computer Science, University of Toronto (2010).

⁴⁴ Geoffrey Hinton: *A Practical Guide to Training Restricted Boltzmann Machines*. En: *Department of Computer Science, University of Toronto (2010).*

⁴⁵ Geoffrey Hinton: *A Practical Guide to Training Restricted Boltzmann Machines*. En: *Department of Computer Science, University of Toronto (2010).*

$$Z = \sum_{\mathbf{x}, \mathbf{n}^{(1)}} e^{-E(\mathbf{x}, \mathbf{n}^{(1)})} \quad (8)$$

⁴⁶ En la Máquina de Boltzmann restringida, la palabra “restringida” se refiere a la estructura del modelo, que es una red neuronal. La red neuronal se divide en dos conjuntos de neuronas y estas no pueden conectarse. En una RBM, las unidades ocultas no pueden interactuar directamente con otras unidades ocultas, y las unidades visibles no pueden interactuar directamente con otras unidades visibles.

4.2. El Algoritmo de K-Medias

⁴⁷ El algoritmo **K-means** es una técnica de agrupamiento iterativa basada en centroides que divide un conjunto de datos en k grupos o conglomerados, con el objetivo de minimizar la distancia entre los puntos y el centro del grupo al que pertenecen.⁴⁸ El centroide, en este contexto, representa la media de los puntos que conforman un clúster. Es decir, un centroide en k-medias es el punto que representa el centro de un grupo de datos, calculado como el promedio de todos los puntos del clúster. Su funcionamiento se basa en una serie de pasos repetitivos: primero, se inicializan aleatoriamente k centroides; luego, en cada iteración, se asigna cada punto al centroide más cercano (utilizando generalmente la distancia euclidiana), se recalculan los centroides como la media de los puntos asignados y se repite este proceso hasta alcanzar la convergencia.

⁴⁹ Para la asignación de clústeres, a cada punto de datos $\mathbf{x}_i \in \mathbb{R}^n$, se calcula su distancia a cada uno de los k centroides $\mu_1, \mu_2, \dots, \mu_k$. Luego, el punto \mathbf{x}_i se asigna al clúster cuyo centroide esté más cercano, es decir, al que tenga la menor distancia euclidiana. Esta asignación se expresa matemáticamente como:

⁴⁶ F. Pedregosa et al.: *Scikit-learn: Machine Learning in Python*. En: *Journal of Machine Learning Research* 12 (2011), págs. 2825-2830.

⁴⁷ J. MacQueen: *Some methods for classification and analysis of multivariate observations*. En: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1 (1967), págs. 281-297.

⁴⁸ A. K. Jain: *Data clustering: 50 years beyond K-means*. En: *Pattern Recognition Letters* 31.8 (2010), págs. 651-666.

⁴⁹ R. Xu y D. Wunsch: *Survey of clustering algorithms*. En: *IEEE Transactions on Neural Networks* 16.3 (2005), págs. 645-678.

$$C_j = \{x_i \mid \|x_i - \mu_j\|^2 \leq \|x_i - \mu_l\|^2, \forall l \in \{1, \dots, k\}\} \quad (9)$$

Esto significa que el punto x_i se incluye en el clúster C_j si la distancia entre x_i y el centroide μ_j es menor o igual que la distancia entre x_i y cualquier otro centroide μ_l .

El centroide $\mu_j \in \mathbb{R}^n$ se define como:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (10)$$

donde:

- $x_i \in \mathbb{R}^n$ es el vector de entrada del punto de datos i ,
- C_j es el conjunto de datos asignados al clúster j ,
- $|C_j|$ representa la cantidad de elementos en el clúster j ,
- μ_j es un vector que contiene la media de cada dimensión para los puntos del clúster.

⁵⁰ La distancia euclidiana al cuadrado es comúnmente utilizada para medir la similitud entre los puntos y los centroides, y se expresa como:

$$d^2(x_i, \mu_j) = \sum_{l=1}^n (x_{il} - \mu_{jl})^2$$

- $x_i = [x_{i1}, x_{i2}, \dots, x_{in}] \in \mathbb{R}^n$: vector correspondiente al i -ésimo dato de entrada.
- $\mu_j = [\mu_{j1}, \mu_{j2}, \dots, \mu_{jn}] \in \mathbb{R}^n$: vector que representa el centroide del clúster C_j .
- n : número de dimensiones (o características) del conjunto de datos.

⁵⁰ R. Xu y D. Wunsch: *Survey of clustering algorithms*. En: *IEEE Transactions on Neural Networks* 16.3 (2005), págs. 645-678.

El objetivo de K-means es minimizar la suma de errores cuadráticos dentro de los clústeres:

$$EC = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2$$

donde

- C_j : conjunto de puntos asignados al clúster j .
- μ_j : centroide del clúster j .
- $\|\cdot\|$: norma euclidiana (distancia euclidiana).

Un valor pequeño de EC indica que las agrupaciones formadas por el algoritmo K-means son de mejor calidad, ya que los puntos están más cercanos a sus respectivos centroides. Esto implica que los clústeres son más compactos y presentan mayor homogeneidad interna.

⁵¹Por último, es importante tener en cuenta que K-means tiene algunas limitaciones. Por ejemplo, es sensible a los datos que se alejan mucho del resto, necesita que se indique de antemano cuántos grupos se quieren formar, estas características hacen necesaria la aplicación de métodos complementarios de validación de clústeres y técnicas de optimización.

4.3. Entropía

En este proyecto, la entropía se utiliza para entender cuánta información o incertidumbre hay en los resultados que genera la máquina de Boltzmann. Esta medida permite observar si los datos, luego de ser transformados por la máquina de Boltzmann, están más organizados o dispersos. Además, compara cómo cambia esa incertidumbre en los años 2020, 2021 y 2023.

⁵² El entrenamiento de la máquina de Boltzmann se considera el impulso de un sistema desde un estado de equilibrio inicial al estado de equilibrio objetivo al cambiar los parámetros del modelo. Es interesante observar cómo cambian las cantidades termodinámicas como la energía libre, la entropía, la energía interna

⁵¹ A. K. Jain y R. C. Dubes: *Algorithms for Clustering Data*. Prentice-Hall, 1988.

⁵² G. Torlai y R. G. Melko: *Learning thermodynamics with Boltzmann machines*. En: *Entropy* 22.5 (2020), pág. 538. DOI: 10.3390/e22050538. <https://www.mdpi.com/1099-4300/22/5/538>.

y el trabajo a medida que avanza el entrenamiento.

- ⁵³ **Función de pérdida:** La función de pérdida, también conocida como función de error, es esencial en el aprendizaje automático, ya que cuantifica la diferencia entre los resultados previstos de un algoritmo de machine learning y los valores objetivo reales, es decir, evalúan la diferencia entre las predicciones de un modelo y los valores reales del conjunto de datos. Estas funciones se utilizan para optimizar el rendimiento del modelo minimizando el error de predicción. Después de calcular la función de pérdida, que está determinada por el error de predicción, el algoritmo de aprendizaje ajusta los pesos y parámetros del modelo para reducir aún más el error en las siguientes iteraciones de entrenamiento.

Existen distintos tipos de funciones de pérdida según las tareas para las que se utilizan. La mayoría de estas funciones están diseñadas para problemas de regresión y clasificación.

- ⁵⁴ **Función de pérdida de regresión:** miden los errores en las predicciones que implican valores continuos. Aunque se aplican de forma más intuitiva a los modelos que estiman directamente conceptos cuantificables como el precio, la antigüedad, el tamaño o el tiempo, la regresión tiene una amplia gama de aplicaciones. Por ejemplo, una función de pérdida de regresión puede utilizarse para optimizar un modelo de imagen cuya tarea consiste en estimar el valor de color de los píxeles individuales.
- ⁵⁵ **Funcion pérdida de entropía cruzada binaria:** La pérdida binaria de entropía cruzada (BCE) es un indicador fundamental para evaluar el desempeño de modelos de clasificación. En este contexto, el modelo genera predicciones como valores de probabilidad entre 0 y 1. Cada valor representa la probabilidad de que una instancia de datos pertenezca a una clase específica. La ecuación matemática de la pérdida binaria de entropía cruzada, también conocida como pérdida logarítmica, es:

⁵³ *Explicación de las funciones de pérdida en el machine learning.* En: *Datacamp* (2024). <https://www.datacamp.com/es/tutorial/loss-function-in-machine-learning>.

⁵⁴ *Explicación de las funciones de pérdida en el machine learning.* En: *Datacamp* (2024). <https://www.datacamp.com/es/tutorial/loss-function-in-machine-learning>.

⁵⁵ *Explicación de las funciones de pérdida en el machine learning.* En: *Datacamp* (2024). <https://www.datacamp.com/es/tutorial/loss-function-in-machine-learning>.

$$\text{Entropía}(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i))$$

siendo y_i para una red binaria 0 o 1, y \hat{y}_i representa la predicción del modelo. La ecuación anterior se aplica específicamente a un escenario en el que el algoritmo de machine learning hará una clasificación entre dos clases. Se trata de un escenario de clasificación binaria.

- ⁵⁶ **Funcion pérdida de entropía cruzada generalizada:** La función perdida de entropía cruzada generalizada se aplica a escenarios de clasificación multiclase. Para K clases que la ecuación corresponde a:

$$\text{Entropía}_G(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n y_i \ln(\hat{y}_i)$$

con $y_i = 0, 1, \dots, K$. Cuando y_i no toma valores numéricos, entonces se toma $p_i = P(Y = y_i)$ en vez de y_i .

Para establecer un grado de comparabilidad de la entropía se suele tomar esta por

$$\text{Entropía}_G(p_i, \hat{p}_i) = -\frac{1}{n} \sum_{i=1}^n p_i \ln(\hat{p}_i)$$

Cuyo valor máximo es 1 cuando $p_i = 0,5$. Cuando se desconocen los valores de p_i , se toma como la entropía la función:

$$\text{Entropía}_G(\hat{p}_i) = -\frac{1}{n} \sum_{i=1}^n \hat{p}_i \ln(\hat{p}_i)$$

La entropía se interpreta como la incertidumbre que existe en cualquier experimento o señal aleatoria. También representa la cantidad de ruido o desorden que tiene un modelo.

Definición 4.3.1. ⁵⁷ La **energía libre** asociada a un vector de entrada \mathbf{x} se define como:

⁵⁶ IBM: ¿Qué es una función de pérdida? En: *IBM Think* (2023). <https://www.ibm.com/es-es/think/topics/loss-function>.

⁵⁷ Giulio Lopez et al.: *Entropy, Free Energy, and Work of Restricted Boltzmann Machines*. En: *Entropy* 22.5 (2020), pág. 538. DOI: 10.3390/e22050538. <https://www.mdpi.com/1099-4300/22/5/538>.

$$F(x) = -\log \sum_{n^{(1)}} p(x, n^{(1)}) = \frac{1}{Z} \sum_{n^{(1)}} e^{-E(x, n^{(1)})} \quad (11)$$

podemos ver que:

$$p(x) = \frac{1}{Z} e^{-F(x)} \quad (12)$$

lo que muestra la relación directa entre la energía libre y la probabilidad de un vector de entrada.

Definición 4.3.2. ⁵⁸ **Entropía:** La entropía de la distribución conjunta $p(x, n^{(1)})$ está definida como:

$$S = - \sum_{x, n^{(1)}} p(x, n^{(1)}) \log p(x, n^{(1)}) \quad (13)$$

Usando que:

$$p(x, n^{(1)}) = \frac{1}{Z} e^{-E(x, n^{(1)})} \quad (14)$$

se puede expresar la entropía como:

$$S = \mathbb{E}_p[E(x, n^{(1)})] + \log Z \quad (15)$$

donde \mathbb{E}_p denota la esperanza con respecto a la distribución $p(x, n^{(1)})$. Note que $\log Z$ actúa como un término de normalización global (dado que Z es la suma de todos los pesos de energía posibles, ver Ecuación 1.5). Finalmente, la entropía mide la “incertidumbre” de la distribución de probabilidad aprendida por la máquina de Boltzmann lo cual permite saber si el modelo está “seguro” al identificar patrones socioeconómicos que influyen en la muerte fetal. Una entropía más baja luego del entrenamiento sugiere que la red ha aprendido patrones más definidos y útiles.

Ejemplo: Supongamos una máquina de Boltzmann con:

⁵⁸ Giulio Lopez et al.: *Entropy, Free Energy, and Work of Restricted Boltzmann Machines*. En: *Entropy* 22.5 (2020), pág. 538. DOI: 10.3390/e22050538. <https://www.mdpi.com/1099-4300/22/5/538>.

- 2 unidades visibles: $x = (x_1, x_2)$
- 2 unidades ocultas: $n^{(1)} = (h_1, h_2)$
- Pesos sinápticos:
- Sesgos: $b_i = 0$

$$B = \begin{bmatrix} 1,0 & -1,20,5 & 0,8 \end{bmatrix}$$

Y tomamos una configuración binaria:

$$x = (1, 0), \quad h = (1, 1)$$

Entonces la energía es:

$$E(x, h) = - \sum_{i,j} B_{ij} x_i h_j = -(1,0 \cdot 1 \cdot 1 + (-1,2) \cdot 1 \cdot 1 + 0,5 \cdot 0 \cdot 1 + 0,8 \cdot 0 \cdot 1) = -(-0,2) = 0,2$$

Luego, la probabilidad se obtiene por:

$$p(x, h) = \frac{1}{Z} \exp(-E(x, h)) = \frac{1}{Z} \exp(-0,2)$$

Este mismo cálculo se realiza para todas las combinaciones posibles de x y h , y luego se normaliza con:

$$Z = \sum_{x,h} \exp(-E(x, h))$$

Definición 4.3.3. Entropía de Shannon⁵⁹ Sea $\mathbf{p} = (p_1, p_2, \dots, p_n)$ una distribución de probabilidad discreta, donde $p_k \geq 0$ para todo k y $\sum_{k=1}^n p_k = 1$. La **entropía de Shannon** se define como:

$$H(\mathbf{p}) = - \sum_{k=1}^n p_k \log_b p_k$$

⁵⁹ Daniel A. Moreira y Mariana A. M. de Souza: Teorema de Khinchin: Uma justificativa para a fórmula da entropia de Shannon. En: *Revista Brasileira de História da Matemática* 20.42 (2020), págs. 1-17. DOI: 10.36908/rbhm.v20i42.337. <https://rbhm.org.br/index.php/RBHM/article/view/337/318>.

donde b es la base del logaritmo (comúnmente $b = 2$ para expresar la entropía en bits). Esta fórmula mide la incertidumbre asociada a la distribución p . Una entropía más alta indica mayor dispersión o desorden; una más baja, mayor concentración o certeza.

Para lograr un aprendizaje efectivo en Machine Learning, la elección de la función de pérdida adecuada es crucial, ya que esta guía el proceso de ajuste de los parámetros del modelo minimizando el error. Se emplea la entropía de Shannon como medida para evaluar la homogeneidad dentro de los clústeres generados mediante técnicas no supervisadas.

Inicialmente, se presentaron las funciones de pérdida de entropía cruzada binaria y entropía cruzada generalizada, ampliamente utilizadas en tareas de clasificación supervisada, con el fin de contextualizar las distintas formas en que la entropía puede aplicarse en distintos enfoques. Sin embargo, en el presente trabajo, se adopta la entropía de Shannon en un contexto no supervisado para medir la incertidumbre asociada a variables categóricas dentro de cada agrupación.

5. CASOS DE ESTUDIO: FACTORES SOCIOECONÓMICOS ASOCIADOS A MUERTES FETALES.

5.1. Descripción de la base de datos.

La base de datos de estadísticas vitales del DANE (Departamento Administrativo Nacional de Estadística), sección "Demografía y población", se utilizará para este trabajo de grado. Los datos corresponden a los años 2020, 2021 y 2023 son de libre acceso. Se seleccionaron los registros de defunciones fetales excluyendo nacimientos y defunciones no fetales. La información proviene de registros administrativos oficiales reportados en todo el territorio nacional. Dado que se incluyen todos los eventos notificados, la fuente de datos corresponde a un sistema de tipo censal, sin aplicación de muestreo probabilístico. Cada registro contiene información detallada sobre el evento de defunción, incluyendo variables sociodemográficas, clínicas y geográficas de la madre y del evento fetal. Esta estructura permite analizar relaciones entre condiciones sociales y la ocurrencia de muertes fetales.

Se utilizó la Máquina de Boltzmann Restringida como etapa inicial del proyecto, con el fin de obtener representaciones a partir de las variables seleccionadas para reducir la complejidad de los datos, eliminar redundancias y facilitar el descubrimiento de patrones ocultos en las condiciones socioeconómicas asociadas a las muertes fetales. Estas representaciones corresponden a nodos en la capa oculta de la máquina de Boltzmann, los cuales no son variables observables, sino características aprendidas por el modelo no supervisado. Estas características se representan como vectores numéricos abstractos, que no son comprensibles por sí mismos. Por esta razón, es necesario aplicar un algoritmo de agrupamiento como K-medias, que facilita su interpretación. Por ejemplo, un nodo podría activarse al detectar simultáneamente condiciones como bajo nivel educativo, ausencia de seguridad social y residencia en un departamento vulnerable. Este tipo de comportamiento ocurre internamente en la red, y puede ser interpretado externamente a través del análisis de las agrupaciones formadas por K-medias.

Estas características se expresan como vectores numéricos abstractos de tamaño 10, ya que la máquina de Boltzmann Restringida fue configurada con 10 nodos ocultos para que pudiera generar representaciones compactas de los datos, se eligió como punto de partida para equilibrar la capacidad de representación y simplicidad del modelo. Más adelante, en el análisis se visualiza como esta elección permite capturar patrones relevantes en los datos. Posteriormente, se aplicó K-medias como algoritmo de agrupamiento

no supervisado, con el objetivo de clasificar los vectores en grupos con características socioeconómicas similares. Se implementó en dos etapas: **antes** y **después** de la máquina de Boltzmann, la primera utilizando directamente las variables originales codificadas y la segunda, utilizando las representaciones generadas por la red, esta doble aplicación se realizó con el fin de comparar los resultados y visualizar el valor que aporta la red en términos de organización y estructuración de los datos. Además, se calculó la entropía para cada etapa con el fin de analizar el nivel de organización y predictibilidad en las representaciones obtenidas, esto permite cuantificar la diversidad o dispersión de las categorías socioeconómicas (como nivel educativo, edad de la madre, departamento o seguridad social). Aplicar la entropía en ambas fases es importante, ya que permite observar la homogeneidad o lo caótico que son los clústeres, dependiendo de la fase (si son los datos originales o las representaciones aprendidas por la máquina de Boltzmann) es decir, al comparar las agrupaciones generadas por las fases mencionadas anteriormente, se puede observar cuál de las dos representaciones genera clústeres más homogéneos y, por tanto, más interpretables. Esta información podría utilizarse para desarrollar intervenciones para reducir la tasa de mortalidad fetal. Es importante resaltar que las variables de la base de datos son cualitativas, pero se representan de forma cuantitativa mediante el algoritmo **One-Hot Encoding**, este preprocesamiento es necesario ya que los algoritmos utilizados, requieren datos numéricos para operar correctamente, en concreto variables binarias, debido a que permite representar la información categórica para facilitar su análisis.

Nombre de la variable	Variable	Tipo de variable	Valores de la variable
Departamento donde ocurrió la defunción	COD_DPTO	Categórica nominal	Códigos numéricos para departamentos (05: Antioquia, 08: Atlántico, etc.)
Municipio donde ocurrió la defunción	COD_MUNIC	Categórica nominal	Códigos DANE para municipios
Área donde ocurrió la defunción	A_DEFUN	Categórica nominal	1: Cabecera, 2: Centro poblado, 3: Rural disperso

Nombre de la variable	Variable	Tipo de variable	Valores de la variable
Sitio donde ocurrió la defunción	SIT_DEFUN	Categórica nominal	1: Institución de salud, 2: Vivienda, 3: Vía pública, 4: Otro sitio
Otro sitio donde ocurrió la defunción	OTRSITIODE	Categórica nominal	Texto libre si se selecciona "Otro sitio"
Tipo de defunción	TIPO_DEFUN	Categórica nominal	1: Fetal, 2: Menor de un año
Año en que ocurrió la defunción	ANO	Cuantitativa discreta	Ej. 2022
Mes en que ocurrió la defunción	MES	Cuantitativa discreta	1 a 12 (enero a diciembre)
Hora en que ocurrió la defunción	HORA	Cuantitativa discreta	0 a 23
Minutos en que ocurrió la defunción	MINUTOS	Cuantitativa discreta	0 a 59
Sexo del fallecido	SEXO	Categórica nominal	1: Masculino, 2: Femenino
País de residencia habitual del fallecido	CODPRES	Categórica nominal	Códigos de país, si no es Colombia
Departamento de residencia habitual	CODPTORE	Categórica nominal	Códigos numéricos para departamentos
Municipio de residencia habitual	CODMUNRE	Categórica nominal	Códigos DANE para municipios

Nombre de la variable	Variable	Tipo de variable	Valores de la variable
Área de residencia habitual	AREA_RES	Categórica nominal	1: Cabecera, 2: Centro poblado, 3: Rural disperso
Régimen de seguridad social	SEG_SOCIAL	Categórica nominal	1: Contributivo, 2: Subsidiado, 3: Especial, 4: Ninguno
Entidad Administradora en Salud	IDADMISALUD	Categórica nominal	Códigos de EPS / IPS según DANE
Probable manera de muerte	P_PMAN_IRIS	Categórica nominal	1: Natural, 2: Accidental, 3: Suicidio, 4: Homicidio, 5: No determinada
Certificado de defunción expedido por	CONS_EXP	Categórica nominal	1: Médico, 2: Autoridad judicial
Muerte relacionada con el parto	MU_PARTO	Categórica nominal	1: Durante el parto, 2: Después del parto
Tipo de parto	T_PARTO	Categórica nominal	1: Vaginal, 2: Cesárea
Multiplicidad del embarazo	TIPO_EMB	Categórica nominal	1: Único, 2: Múltiple
Semanas de gestación	T_GES	Cuantitativa discreta	Número de semanas
Peso al nacer	PESO_NAC	Cuantitativa continua	Peso en gramos
Edad de la madre	EDAD_MADRE	Cuantitativa discreta	Edad en años

Nombre de la variable	Variable	Tipo de variable	Valores de la variable
Número de hijos nacidos vivos	N_HIJOSV	Cuantitativa discreta	Número total incluyendo el presente
Número de hijos nacidos muertos	N_HIJOSM	Cuantitativa discreta	Número total incluyendo el presente
Estado conyugal de la madre	EST_CIVM	Categórica nominal	1: Soltera, 2: Casada, 3: Unión libre, etc.
Nivel de estudios de la madre	NIV_EDUM	Categórica ordinal	1: Primaria, 2: Secundaria, 3: Técnica, 4: Universitaria
Último año o grado aprobado	ULTCURMAD	Cuantitativa discreta	Número del grado o año aprobado
Departamento del hecho violento	CODOCUR	Categórica nominal	Códigos numéricos para departamentos
Municipio del hecho violento	CODMUNOC	Categórica nominal	Códigos DANE para municipios
Causa de muerte A	C_MUERTE	Categórica nominal	Códigos CIE-10
Causa de muerte B	C_MUERTEB	Categórica nominal	Códigos CIE-10
Causa de muerte C	C_MUERTEC	Categórica nominal	Códigos CIE-10
Causa de muerte D	C_MUERTE D	Categórica nominal	Códigos CIE-10
Causa de muerte E	C_MUERTEE	Categórica nominal	Códigos CIE-10

Nombre de la variable	Variable	Tipo de variable	Valores de la variable
Asistencia médica antes de fallecer	ASIS_MED	Categórica nominal	1: Sí, 2: No
Causas múltiples antecedentes	CAUSA_MULT	Categórica nominal	Códigos múltiples CIE-10
Causa básica de defunción	C_BAS1	Categórica nominal	Código CIE-10 principal
Causa agrupada OPS	CAUSA_667	Categórica nominal	Categoría según lista 6/67 OPS
Profesión del certificador	IDPROFCER	Categórica nominal	1: Médico, 2: Enfermero, etc.
Causa básica Colombia	CAU_HOMOL	Categórica nominal	Agrupación según Lista 105
Agrupación de semanas de gestación	T_GES_AGRU_CIE	Categórica ordinal	1: <22 sem, 2: 22–27, 3: 28–36, 4: 37+

Tabla 2. Descripción de las variables socioeconómicas, demográficas y médicas utilizadas en el análisis de defunciones fetales en Colombia.

Las variables de interés para la realización de este proyecto son el departamento donde ocurrió la defunción, el régimen de seguridad social de la madre, el nivel educativo y la edad de la madre (véase el Anexo A: Variables de interés). En este caso, el número de defunciones fetales en los años 2020, 2021 y 2023 fue de 33.327, 30.709 y 24.682 (ver⁶⁰).

⁶⁰ Dane, estadísticas vitales: *Defunciones fetales y no fetales. 3er trimestre 2023pr, año corrido 2023pr*,

5.2. Análisis de factores socioeconómicos asociados a defunciones fetales.

- **Limpieza de la base de datos 2020.** Se realizó la limpieza para las cuatro variables seleccionadas: `seg_social` (seguridad social), `niv_edum` (nivel educativo) y `edad_madre` (edad de la madre), teniendo en cuenta los valores utilizados por el DANE para representar datos faltantes, como el 9 en la variable de seguridad social, el 99 en nivel educativo y el 99 en edad de la madre, se desarrolló un código para depurar estos registros sin información:

```
datos_rangos = datos_importantes[
    (datos_importantes["seg_social"] < 9) &
    (datos_importantes["niv_edum"] < 99) &
    (datos_importantes["edad_madre"] < 99)
]

total_original = len(datos_importantes)
total_filtrado = len(datos_rangos)
porcentaje_eliminado = (total_original - total_filtrado) / total_original * 100
print(f"Porcentaje de datos eliminados: {porcentaje_eliminado:.2f}%")
```

Porcentaje de datos eliminados: 7.71%

La variable correspondiente al código del departamento no presentó registros con datos faltantes, por lo tanto, no fue necesario aplicar un filtro sobre ella. En total, se excluyó el 7,71 % de los registros, correspondiente a casos con información no válida. Esta limpieza fue necesaria para garantizar la consistencia del análisis, ya que el aprendizaje no supervisado requiere datos válidos.

cifras definitivas 2022. En: DANE (2023). <https://www.dane.gov.co/files/operaciones/EEVV/bol-EEVV-Defunciones-IIItrim2023.pdf>.

- **Transformación a variables binarias.**

Antes de aplicar la RBM, se implementó utilizando la clase `OneHotEncoder` de la biblioteca `scikit-learn`, como se muestra en el siguiente fragmento de código:

```
from sklearn.preprocessing import OneHotEncoder

# Aplicar One-Hot Encoding a todas las variables categóricas
encoder = OneHotEncoder(sparse_output=False)
data_encoded = encoder.fit_transform(datos_filtros)
```

El conjunto de datos original contenía solamente variables categóricas, como el nivel educativo, la edad de la madre (agrupada), el departamento y el tipo de seguridad social. Por esta razón, se aplicó **One-Hot Encoding**⁶¹, una técnica que transforma cada categoría en una variable binaria, en este caso son variables Dummies, ya que cada nueva variable representa una categoría única de una variable original, codificada como 1 si la condición esta presente y 0 si no lo está. Como resultado de esta transformación, se generó un total de 60 variables binarias, que representan todas las categorías posibles combinadas de las variables seleccionadas.

Este mismo procedimiento se aplicó a las bases de datos correspondientes a los años 2021 y 2023.

5.2.1. Máquina de Boltzmann Restringida

5.2.1.1. RBM para la base de datos Se aplicó la Máquina de Boltzmann Restringida utilizando la clase `BernoulliRBM` de `scikit-learn`, como se muestra en el siguiente fragmento de código:

⁶¹ F. Pedregosa et al.: *Scikit-learn: Machine Learning in Python*. En: *Journal of Machine Learning Research* 12 (2011), págs. 2825-2830.

```

# Configuración de la Máquina de Boltzmann Restringida (RBM)
rbm = BernoulliRBM(n_components=10, learning_rate=0.01, n_iter=100, random_state=42)

# Entrenar la RBM
datos_rbm = rbm.fit_transform(datos_normalizados)

# Convertir a DataFrame para visualización
datos_rbm_df = pd.DataFrame(datos_rbm)

# Mostrar las primeras filas de las características aprendidas por la RBM
print("Representaciones aprendidas por la RBM:")
print(datos_rbm_df.head())

```

La máquina de Boltzmann (RBM) transformó las variables socioeconómicas originales (categorizadas en variables tipo Dummie con valores de 0 y 1) en nuevas representaciones más compactas, llamadas características latentes. Estas representaciones son vectores de dimensión 10 y corresponden a las combinaciones de las variables originales, que no son binarias estrictas, sino valores continuos dentro del intervalo $[0, 1]$, ya que la salida de cada nodo corresponde a la probabilidad de activación, no a una decisión binaria definitiva. Esta probabilidad se calcula mediante la función de activación sigmoide, que devuelve valores continuos entre 0 y 1, se interpretan como los grados de activación de cada nodo oculto ante el patrón de entrada. Para interpretar esta nueva información, se utilizó el algoritmo K-medias.

Interpretación de las representaciones aprendidas por la RBM para 2020

↔ Representaciones aprendidas por la RBM:

	0	1	2	3	4	5	6	\
0	0.050589	0.993042	0.000273	0.000653	0.999910	0.999922	0.891045	
1	0.001550	0.999194	0.999298	0.999818	0.999920	0.000160	0.005129	
2	0.002357	0.986859	0.999952	0.999609	0.992370	0.999589	0.000004	
3	0.020769	0.843205	0.999975	0.000232	0.003450	0.999976	0.013183	
4	0.990529	0.412500	0.999156	0.996726	0.999936	0.997619	0.000014	
	7	8	9					
0	0.000382	0.997648	0.999880					
1	0.993660	0.979810	0.999423					
2	0.997922	0.947139	0.000017					
3	0.997822	0.993814	0.005542					
4	0.997856	0.243472	0.999995					

Figura 15. Matriz de características latentes para la base de datos del año 2020.

Teniendo en cuenta la información anteriormente mencionada de la RBM, donde trabaja con entradas binarias, las salidas de la capa oculta no son binarias estrictas, si no que corresponden a probabilidades de activación de cada nodo oculto frente al patrón de entrada, las representaciones aprendidas están en el intervalo $[0, 1]$, donde un valor cercano a 1 indica una alta probabilidad de activación del nodo, y un valor cercano a 0 sugiere una baja probabilidad de activación.

Se pudo observar que la mayoría de las representaciones de los datos de 2020 hecha por la RBM, tienen un patrón muy cercano a 1 como se puede observar en las componentes 1, 2, 7, 8, esto sugiere que esos nodos están respondiendo a patrones similares en los datos y las demás componentes son más dispersas porque son más cercanas a 0, lo que indica una baja activación. Sin embargo, es importante señalar que no es posible afirmar directamente un significado socioeconómico específico a cada componente. Para poder afirmar, por ejemplo, que una componente está relacionada con madres de alto nivel educativo o determinada seguridad social, es necesario realizar un análisis complementario, como K-medias.

Interpretación de las representaciones aprendidas por la RBM para 2021

↔ Representaciones aprendidas por la RBM:

	0	1	2	3	4	5	6	\
0	0.742308	0.993090	0.003669	0.987690	2.786126e-07	0.999733	0.007208	
1	0.495721	0.992600	0.014784	0.996498	9.958778e-01	0.994582	0.001193	
2	0.067860	0.991685	0.001329	0.997490	9.998934e-01	0.999040	0.001603	
3	0.004683	0.996543	0.999092	0.999950	9.967056e-01	0.001076	0.996839	
4	0.003684	0.997247	0.999244	0.999504	9.971563e-01	0.012985	0.000800	
	7	8	9					
0	0.000908	0.941193	0.999400					
1	0.000946	0.995714	0.000021					
2	0.002199	0.994808	0.999831					
3	0.979693	0.011903	0.000008					
4	0.997487	0.998658	0.000016					

Figura 16. Matriz de características latentes para la base de datos del año 2021.

Teniendo en cuenta la información anteriormente mencionada de la RBM, donde solo acepta valores binarios, las representaciones aprendidas están en el rango $[0, 1]$, donde valores cercanos a 1 indica que el nodo está activo, mientras que valores cercanos a cero lo contrario.

Para el año 2021, se puede evidenciar que varios componentes tienen valores cercanos a 1, como las columnas 1, 3, 5, 8, lo que indica una fuerte correlación de las representaciones. Caso contrario a las columnas restantes que tienen valores cercanos a 0. Se puede apreciar la diferencia con la base del año 2020, pues en esta las representaciones poco relacionadas eran pocas, a diferencia del nodo 4 del año 2021 que tiene filas muy cercanas a 0.

Interpretación de las representaciones aprendidas por la RBM para 2023

⇒ Representaciones aprendidas por la RBM:

	0	1	2	3	4	5	6	\
0	0.975300	0.976018	0.994173	0.004346	0.997770	0.985276	0.000618	
1	0.972211	0.000125	0.997382	0.000607	0.979433	0.976982	0.000954	
2	0.983446	0.974714	0.994288	0.003553	0.997113	0.988267	0.997600	
3	0.011119	0.185033	0.994583	0.997208	0.000039	0.005997	0.000306	
4	0.955355	0.013780	0.997187	0.001358	0.994058	0.967237	0.998403	
	7	8	9					
0	0.000355	0.032685	0.991810					
1	0.999467	0.355838	0.998397					
2	0.000402	0.004187	0.995932					
3	0.998699	0.963100	0.997577					
4	0.998912	0.000055	0.995710					

Figura 17. Matriz de características latentes para la base de datos del año 2023.

Teniendo en cuenta la información anteriormente mencionada de la RBM, donde solo acepta valores binarios, las representaciones aprendidas están en el rango $[0, 1]$, donde valores cercanos a 1 indica que el nodo está activo, mientras que valores cercanos a cero lo contrario.

Finalmente para el año 2023, se observa que las representaciones están más cercanas a 1, a diferencia de los años anteriores, lo que indica que para este año la RBM encontró más estabilidad en las variables.

5.2.1.2. Error de reconstrucción Se aplicó la función de activación Sigmoide de manera manual:

```
# Calcular la reconstrucción manualmente
def sigmoide(x):
    return 1 / (1 + np.exp(-x))

# Propagar hacia la capa oculta
activaciones_ocultas = sigmoide(np.dot(data_encoded, rbm.components_.T) +
```

```
rbm.intercept_hidden_)

# Reconstruir la entrada desde la capa oculta
X_reconstruido = sigmoide(np.dot(activaciones_ocultas, rbm.components_) +
rbm.intercept_visible_)

# Calcular el error de reconstrucción
error_reconstruccion = np.mean((data_encoded - X_reconstruido) ** 2)
print(f"Error de reconstrucción de la RBM: {error_reconstruccion}")
```

Para visualizar si la RBM tuvo un buen aprendizaje, se realizó el error de reconstrucción mediante la función Sigmoide que es una función de activación, la cual convierte los valores de salida del modelo en probabilidades entre 0 y 1. El error de reconstrucción hace una comparación entre las variables originales y las reconstruidas manualmente; un error bajo (cercano a 0) nos sugiere que la RBM aprendió de manera correcta y las representaciones son apropiadas, mientras que un error cercano a 1 nos indica un sobreajuste de la máquina.

Los errores de reconstrucción obtenidos al aplicar el código para los años 2020, 2021 y 2023 fueron los siguientes:

Error de reconstrucción de la RBM: 0.0156663118801355

Figura 18. Error de reconstrucción de la RBM para el año 2020.

Error de reconstrucción de la RBM: 0.020264900707146455

Figura 19. Error de reconstrucción de la RBM para el año 2021.

Error de reconstrucción de la RBM: 0.01669078641591624

Figura 20. Error de reconstrucción de la RBM para el año 2023.

Se concluye que para el aprendizaje de la RBM en los años seleccionados, se tuvo un buen error de reconstrucción ya que son cercanos a cero.

5.2.2. Matriz de correlación Para hallar la matriz de correlación entre las representaciones de salida de la RBM se implemento Seaborn⁶² la cual es una librería de visualización de datos, que contiene funciones o clases para crear gráficos.

```
import seaborn as sns
import matplotlib.pyplot as plt

# Calculamos la matriz de correlación de las variables después de la RBM
correlation_matrix = pd.DataFrame(datos_rbm).corr()

# Graficamos la matriz de correlación
plt.figure(figsize=(8,6))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Matriz de correlación después de RBM")
plt.show()
```

Luego de la implementación de la RBM, se quiere identificar el grado de dependencia lineal entre las representaciones generadas por la máquina. Al observar esta matriz, es posible identificar variables altamente correlacionadas que podrían estar capturando información similar. Esto proporciona información útil para determinar redundancia o complementariedad entre las nuevas características.

⁶² Michael L. Waskom: *seaborn: statistical data visualization*. En: *Journal of Open Source Software* 6.60 (2021), pág. 3021. DOI: 10.21105/joss.03021. <https://doi.org/10.21105/joss.03021>.

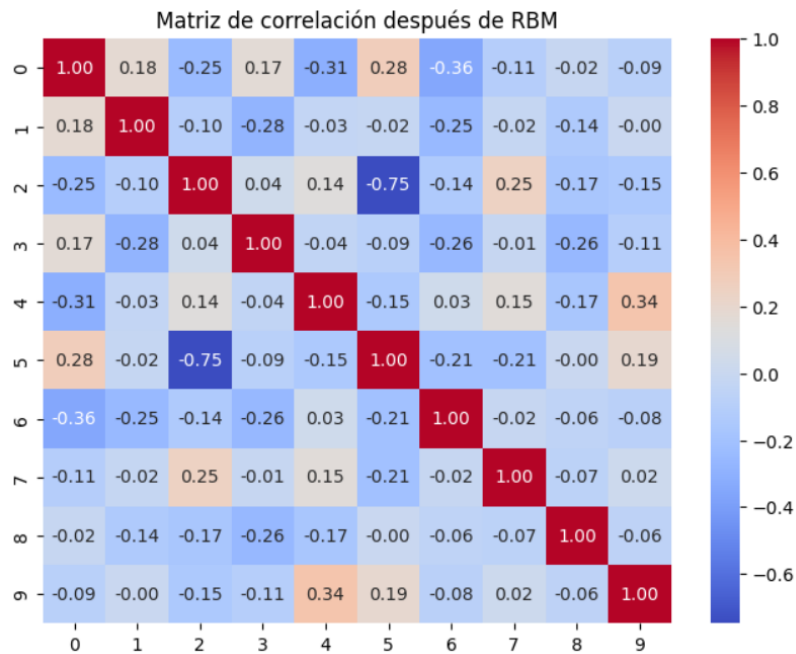


Figura 21. Matriz de correlación después de RBM para el año 2020.

Se apreció que la mayoría de coeficientes en la matriz de correlación están cerca de 0, lo que indica que son independientes entre si. En el caso particular de los nodos 2 y 5 tienen una correlación negativa. También se observó que en el caso de los nodos 4 y 9 es la correlación positiva más alta de la matriz, sin embargo no está cerca de 1 para ser buena.

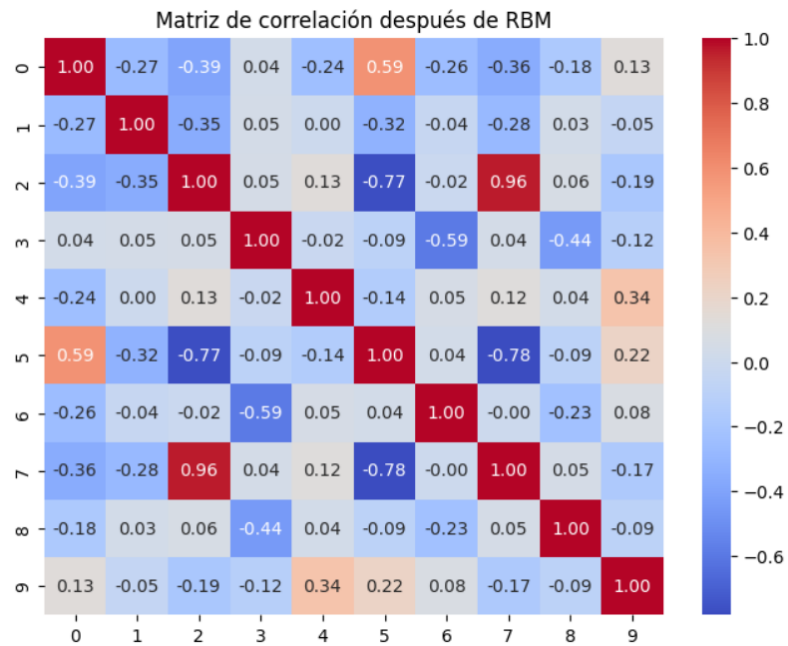


Figura 22. Matriz de correlación después de RBM para el año 2021.

En el año 2021, se observo patrones similares al año anterior, en el caso particular de los nodos 5 y 7 hay una correlación fuertemente negativa y en los nodos 5 y 2 el cual es el valor más cercano a 1, lo que quiere decir que esos nodos están relacionados, y finalmente en los nodos 0 y 3 están cercanos a 0 sugiere, que no están linealmente relacionados.

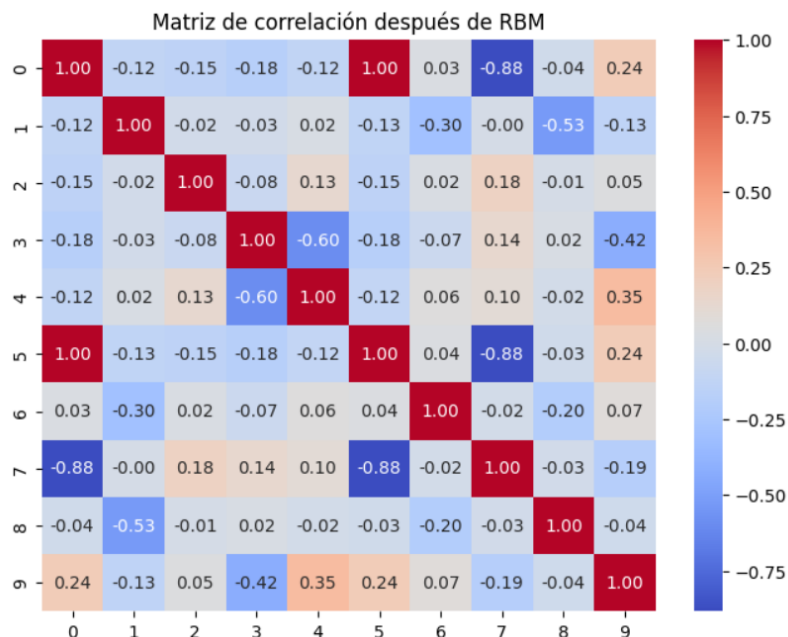


Figura 23. Matriz de correlación después de RBM para el año 2023.

Los resultados indicaron que la mayoría de coeficientes en la matriz de correlación se encuentran cercanos a 0, lo que supone que no hay una relación lineal fuerte entre las variables. En el caso particular de los nodos 0 y 5 tienen una correlación de 1, es decir, capturan información similar o las características de ambos nodos se complementan entre sí. También se observó que en el caso de los nodos 0 y 7 la correlación es altamente negativa, evidencia de que están inversamente relacionados.

Finalmente, se concluye que las variables de las bases de datos de los años seleccionados, representaron en su mayoría poca relación lineal entre los nodos.

5.3. Etapas

Se utilizará el algoritmo K-medias, el cual será implementado en dos etapas: antes y después del entrenamiento de la Máquina de Boltzmann Restringida (RBM). Esta estrategia tiene como propósito evaluar si las representaciones aprendidas por la RBM permiten una mejor estructuración de los datos y, en consecuencia, una segmentación más clara y significativa de

los factores socioeconómicos asociados a las defunciones fetales.

5.3.1. Antes: K-Medias Se aplicó el algoritmo de K-Medias directamente sobre los datos, sin realizar previamente una reducción de dimensionalidad. El propósito fue evaluar si las características eran capaces de identificar agrupamientos significativos en los datos, permitiendo formar grupos con características socioeconómicas similares. Posteriormente, se utilizó el Análisis de Componentes Principales únicamente con fines de visualización, proyectando los clústeres en un espacio bidimensional para facilitar su interpretación gráfica.

```
# Aplicar K-Medias sobre datos normalizados (antes de RBM)
kmeans_preRBM = KMeans(n_clusters=4, random_state=42, n_init='auto')
kmeans_preRBM.fit(datos_normalizados) # Aquí trabajamos con datos_normalizados

# Asignar los clústeres al DataFrame original de variables importantes
datos_importantes["Cluster_preRBM"] = kmeans_preRBM.labels_

# Ver las primeras filas con los clusters asignados
print(datos_importantes.head())
```

Se usó el algoritmo K-Medias⁶³ para dividir los datos en cuatro grupos. Se eligió este número de grupos porque permite hacer una interpretación clara y útil de los resultados. Después, se revisaron los promedios de las variables en cada grupo para entender sus diferencias. Con esto, se intentó identificar qué tipo de grupo social representa cada uno, según sus características.

A continuación, se presenta y se analiza la importancia de las variables en cada clúster en los años 2020, 2021 y 2023, representada en un gráfico tipo *heatmap*.

Año 2020

⁶³ F. Pedregosa et al.: *Scikit-learn: Machine Learning in Python*. En: *Journal of Machine Learning Research* 12 (2011), págs. 2825-2830.

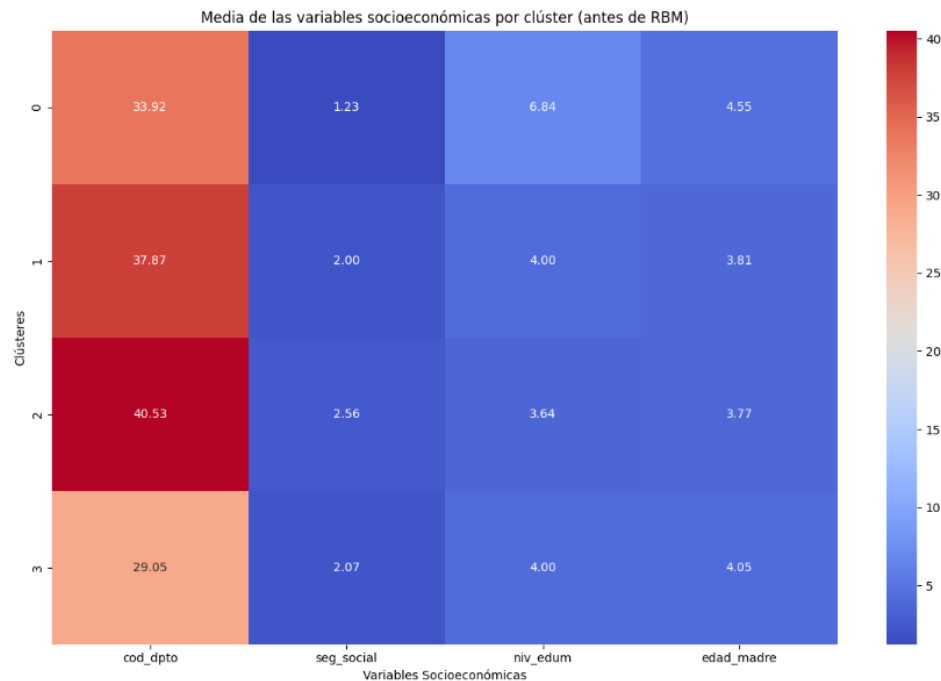


Figura 24. Agrupación por cluster para el año 2020.

Clúster 0:

- Edad de la madre: promedio de 4.55, correspondiente al grupo de 25 a 29 años.
- Nivel educativo: 6.84, lo cual indica predominancia de madres con educación normalista a técnica profesional.
- Régimen de seguridad social: 1.23, lo que refleja afiliación al régimen contributivo en su mayoría.
- Departamento promedio: 33.92, asociado a zonas como Choco o Huila.

Este grupo representa madres adultas jóvenes con buen nivel educativo y afiliadas al regimen contributivo. Están ubicadas en departamentos con condiciones sociales dificiles a intermedias. Se considera una población media vulnerable al riesgo de muerte fetal.

Clúster 1:

- Edad de la madre: 3.81, lo que indica madres entre 20 y 29 años.
- Nivel educativo: 4.00, correspondiente a educación media académica o clásica.
- Régimen de seguridad social: 2.00, lo que implica presencia exclusiva del régimen subsidiado.
- Departamento promedio: 37.87, cercano a departamentos como Chocó o principalmente Huila.

Este clúster refleja una población joven adulta, con educación media y acceso restringido a servicios de salud. La afiliación al régimen subsidiado sugiere condiciones socioeconómicas más limitadas, lo que puede aumentar su vulnerabilidad en el embarazo.

Clúster 2:

- Edad de la madre: 3.77, correspondiente también al rango de 20 a 29 años.
- Nivel educativo: 3.64, predominancia de educación básica secundaria a media académica o clásica.
- Régimen de seguridad social: 2.56, lo que indica una mezcla entre afiliación subsidiada y excepción.
- Departamento promedio: 40.53, lo que sugiere Huila.

Este grupo agrupa a madres jóvenes adultas con bajo nivel educativo y condiciones sociales desfavorables. La elevada codificación de Huila sugiere que es un departamento con mayores condiciones desfavorables. Este perfil se alinea con altos factores de riesgo.

Clúster 3:

- Edad de la madre: 4.05, lo que indica edades entre 25 y 29 años.
- Nivel educativo: 4.00, correspondiente a educación media o clásica.
- Régimen de seguridad social: 2.07, lo que implica predominio del régimen subsidiado.
- Departamento promedio: 29.05, lo cual apunta a zonas como El Choco.

Aunque presenta una edad materna más estable, este clúster comparte similitudes con los clústeres más vulnerables, en cuanto a bajo acceso a servicios de salud y menor nivel educativo.

Para complementar el análisis descriptivo de los clústeres generados por K-means, se presentan a continuación visualizaciones de su distribución en un espacio bidimensional. Estas gráficas permiten evaluar gráficamente la calidad de la agrupación y su relación con las variables socioeconómicas originales.

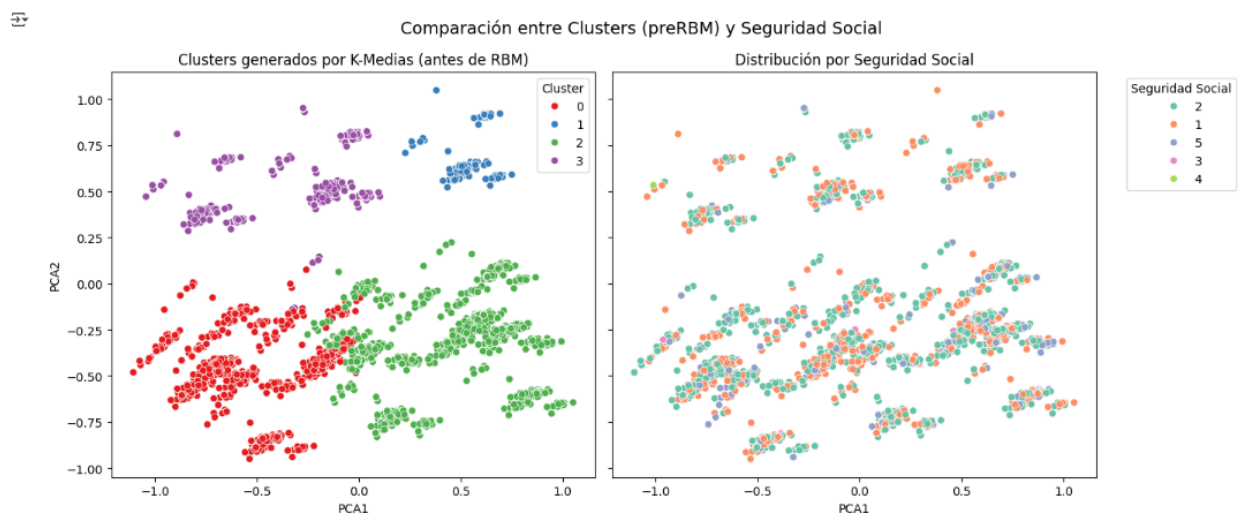


Figura 25. Distribución de clústeres K-means vs Seguridad Social. Año 2020.

Como se puede observar en la Figura 25, los clústeres generados por el algoritmo K-medias antes de aplicar la Máquina de Boltzmann Restringida (pre-RBM) y la variable seguridad social. El clúster 0 muestra una alta concentración de madres afiliadas al régimen contributivo. En contraste, el clúster 2 agrupa a una población mayoritariamente afiliada al régimen subsidiado, lo cual indica

una situación de mayor vulnerabilidad económica. Por su parte, los clústeres 1 y 3 presentan una mezcla más heterogénea, con afiliaciones tanto al régimen subsidiado como a regímenes especiales.

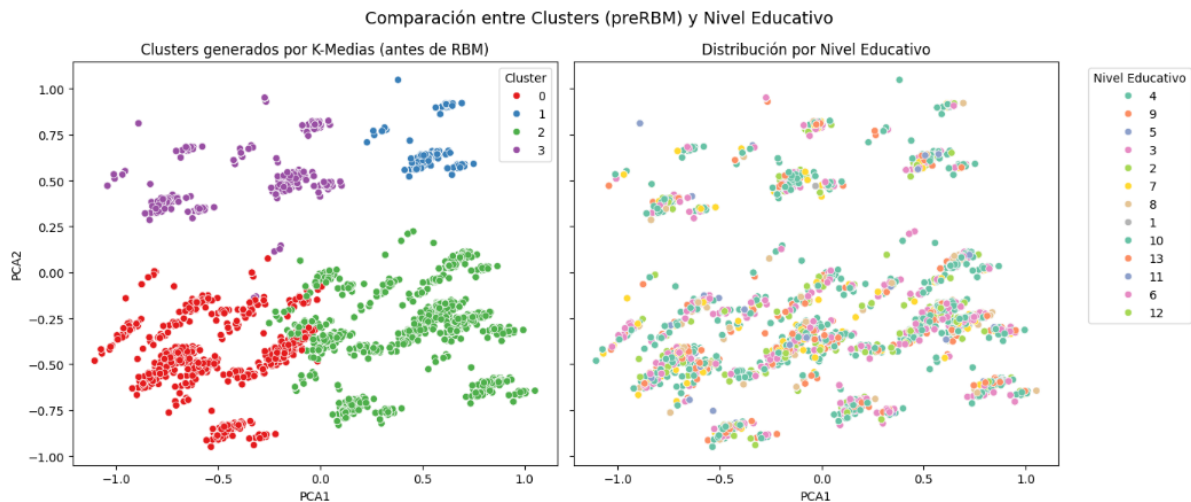


Figura 26. Distribución de clústeres K-means vs Edad Madre. Año 2020.

En la Figura 26 se analiza la relación entre los clústeres obtenidos por K-medias antes de la RBM y la variable nivel educativo. Se observa que el clúster 0 concentra un gran número de madres con educación técnica o profesional no universitaria. Por otro lado, los clústeres 1 y 3 presentan una diversidad de niveles educativos, aunque con predominancia de niveles básicos. El clúster 2 muestra mayor homogeneidad en los niveles de educación media y secundaria.

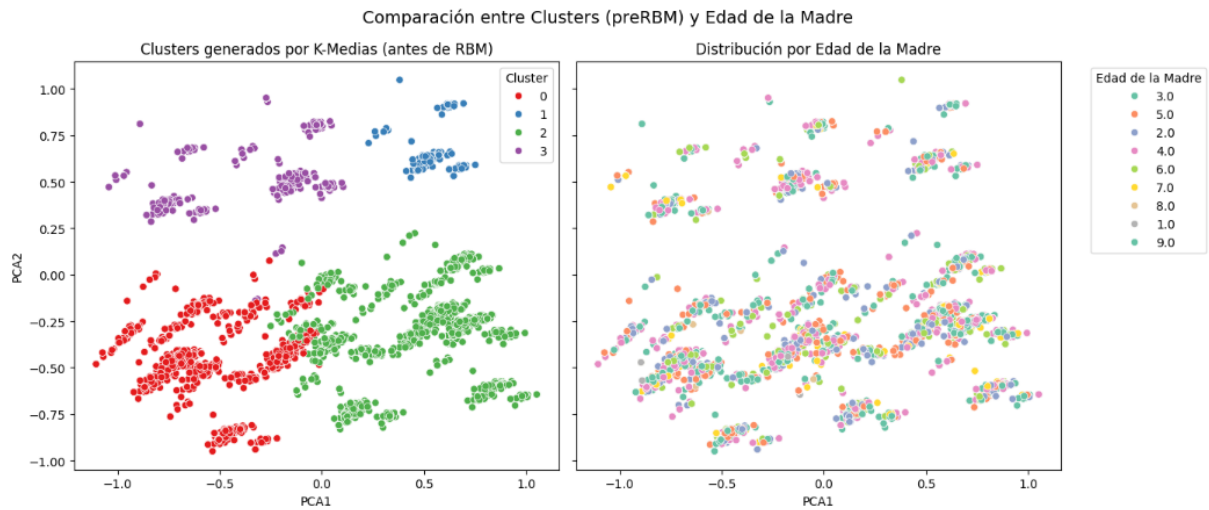


Figura 27. Distribución de clústeres K-means vs Edad Madre. Año 2020.

La Figura 27 muestra la distribución de los clústeres obtenidos por K-medias antes del paso por la RBM frente a la variable edad de la madre. Se observa que todos los clústeres están compuestos principalmente por mujeres entre los 20 y 34 años. No obstante, el clúster 3 presenta una ligera tendencia hacia edades más altas, mientras que el clúster 1 incluye también una fracción relevante de mujeres jóvenes entre 15 y 19 años.

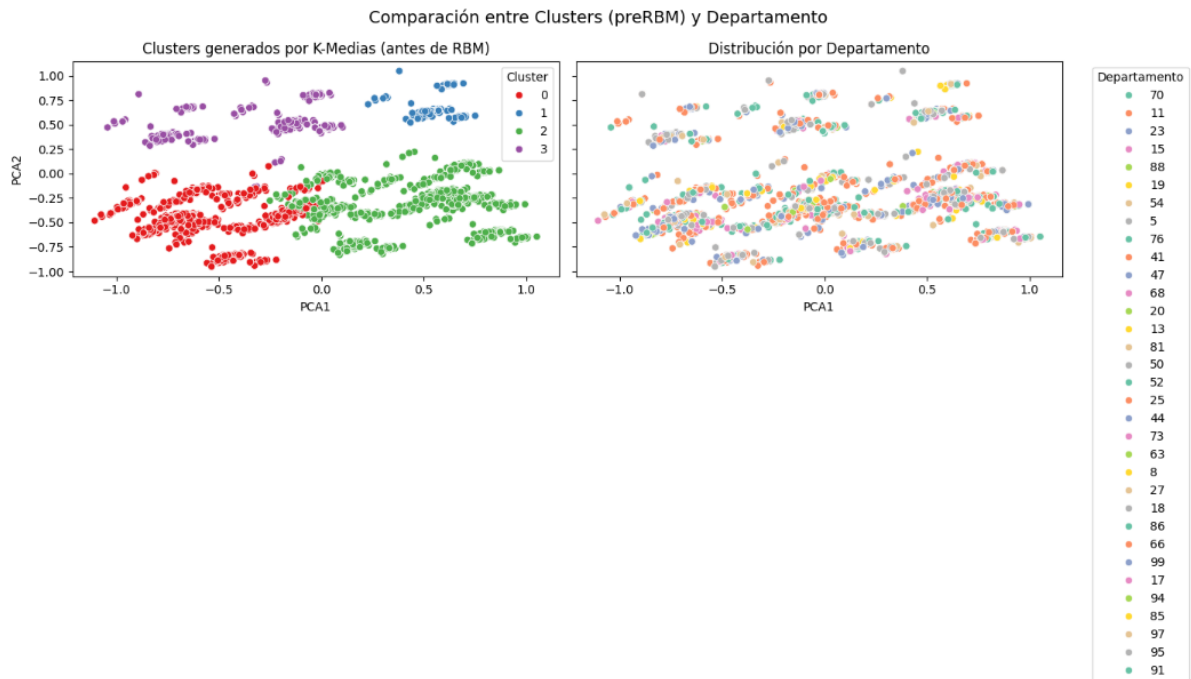


Figura 28. Distribución de clústeres K-means vs Código Departamento. Año 2020.

La figura 28 permite visualizar cómo se distribuyen geográficamente los clústeres obtenidos por el algoritmo K-medias antes de aplicar la RBM. El clúster 0 parece estar compuesto por Córdoba, Cundinamarca y Sucre. El clúster 2, por su parte, agrupa casos provenientes de regiones más vulnerables como Chocó, Magdalena o La Guajira. El clúster 1 presenta una distribución de departamentos intermedios como Bogotá o Antioquia, mientras que el clúster 3 incluye casos de zonas periféricas.

Se realizó el cálculo de la entropía para evaluar la diversidad interna de cada agrupación según las variables.

Entropía por clúster respecto a 'cod_dpto' (antes de RBM):

Clúster 0: Entropía = 3.5244 bits

Clúster 1: Entropía = 4.1108 bits

Clúster 2: Entropía = 4.3868 bits

Clúster 3: Entropía = 3.2295 bits

Entropía por clúster respecto a 'edad_madre' (antes de RBM):

Clúster 0: Entropía = 2.5133 bits

Clúster 1: Entropía = 2.3891 bits

Clúster 2: Entropía = 2.5888 bits

Clúster 3: Entropía = 2.4275 bits

Entropía por clúster respecto a 'niv_edum' (antes de RBM):

Clúster 0: Entropía = 2.4691 bits

Clúster 1: Entropía = 0.0000 bits

Clúster 2: Entropía = 1.9304 bits

Clúster 3: Entropía = 0.0220 bits

Entropía por clúster respecto a 'seg_social' (antes de RBM):

Clúster 0: Entropía = 0.4624 bits

Clúster 1: Entropía = 0.0000 bits

Clúster 2: Entropía = 0.7497 bits

Clúster 3: Entropía = 0.9746 bits

A continuación, se presenta un gráfico de barras que resume la entropía calculada por clúster y por variable, permitiendo identificar visualmente los niveles de homogeneidad o heterogeneidad presentes en cada agrupación.

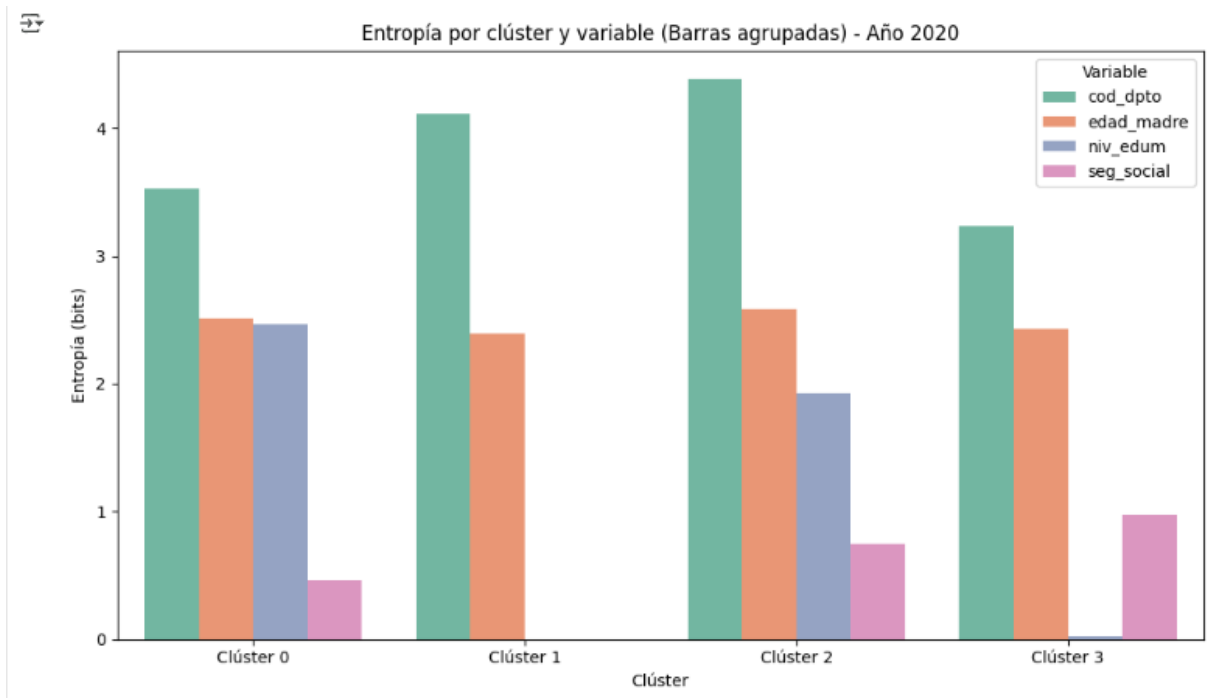


Figura 29. Entropía por clúster y variable socioeconómica. Año 2020.

La Figura 29 muestra que los clústeres 1 y 3 presentan entropía nula en las variables *seguridad social* y *nivel educativo*, respectivamente, lo que indica una homogeneidad total en estas dimensiones dentro de esos grupos. En el caso del clúster 1, todas las observaciones comparten el mismo régimen de afiliación y nivel educativo, mientras que en el clúster 3, la uniformidad se presenta únicamente en la educación. El clúster 0 muestra una mayor diversidad educativa, mientras que el clúster 2 es el más disperso territorialmente, con una alta entropía en la variable *departamento*.

Año 2021

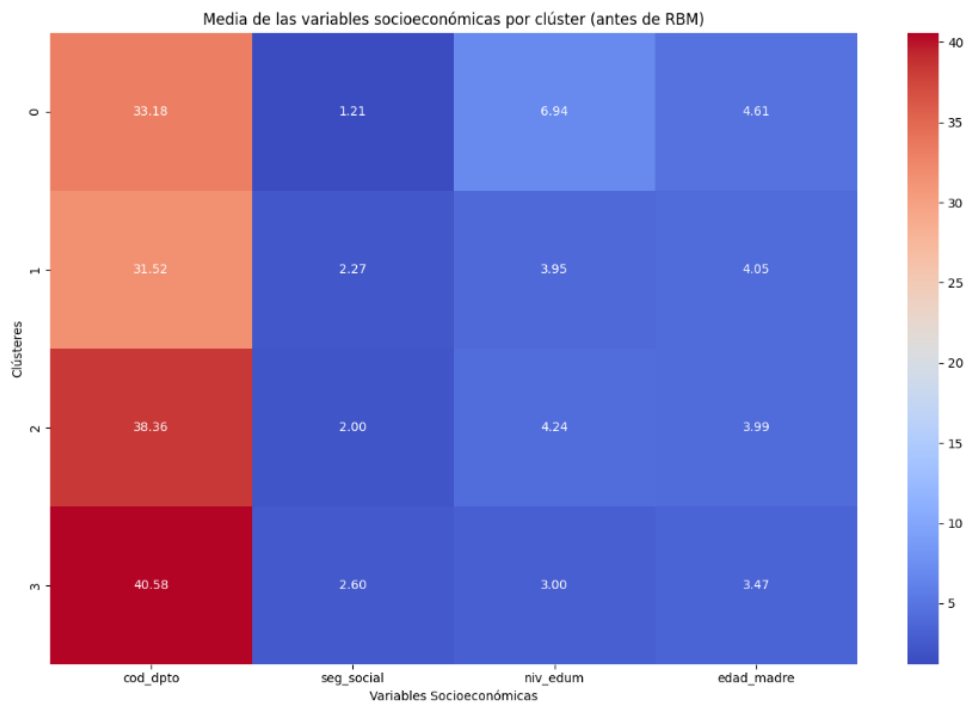


Figura 30. Agrupación por cluster para el año 2021.

Clúster 0:

- Edad de la madre: promedio de 4.61, correspondiente al grupo de 25 a 29 años.
- Nivel educativo: 6.94, lo que indica una concentración en madres con formación normalista a técnica profesional.
- Régimen de seguridad social: 1.21, predominantemente afiliadas al régimen contributivo (EPS).
- Departamento promedio: 33.18, correspondiente a zonas como Choco o Huila.

Este grupo representa madres con condiciones sociales más favorables: mayor nivel educativo, afiliación contributiva y departamentos intermedios. Son madres adultas jóvenes con mejores oportunidades educativas y de salud.

Clúster 1:

- Edad de la madre: 4.05, lo que corresponde a madres entre 25 y 29 años.
- Nivel educativo: 3.95, indicando educación básica secundaria a media académica o clásica.
- Régimen de seguridad social: 2.27, lo que sugiere prevalencia del régimen subsidiado.
- Departamento promedio: 31.52, aproximado a departamentos como Chocó o Huila.

Este grupo muestra un nivel educativo medio y afiliación al régimen subsidiado, por lo que refleja una población con ciertas limitaciones sociales y económicas, aunque no en condiciones extremas.

Clúster 2:

- Edad de la madre: 3.99, lo que corresponde a madres de entre 20 y 29 años.
- Nivel educativo: 4.24, equivalente a educación media académica o clásica.
- Régimen de seguridad social: 2.00, afiliación exclusiva al régimen subsidiado.
- Departamento promedio: 38.36, similar a Huila.

Este grupo presenta características de vulnerabilidad social: afiliación subsidiada y nivel educativo medio-bajo.

Clúster 3:

- Edad de la madre: 3.47, correspondiente a edades entre 20 y 24 años.
- Nivel educativo: 3.00, indicando una predominancia de madres con solo educación básica secundaria.
- Régimen de seguridad social: 2.60, lo que sugiere afiliación subsidiada y posible presencia de excepción (no afiliación).

- Departamento promedio: 40.58, como Huila.

Este grupo refleja bajo nivel educativo, edad joven y afiliación precaria. Además, está localizado nuevamente en Huila, lo que podría indicar mayor vulnerabilidad en ese departamento.

Las gráficas a continuación permiten evaluar gráficamente la calidad de la agrupación y su relación con las variables socioeconómicas originales.

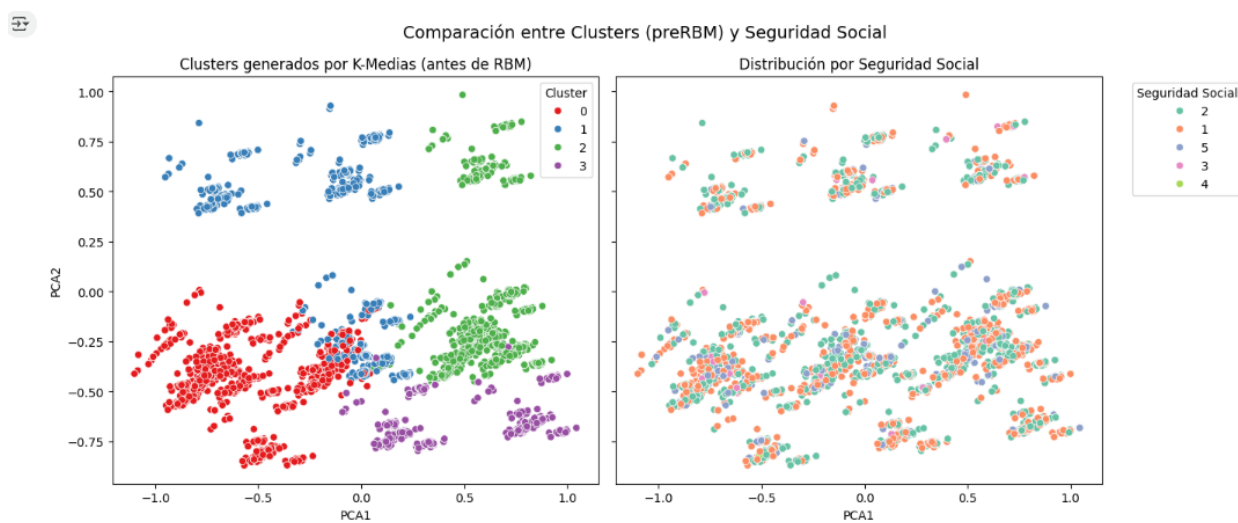


Figura 31. Distribución de clústeres K-means vs Seguridad Social. Año 2021.

Como se puede observar en la Figura 31, los clústeres generados por el algoritmo K-medias antes de aplicar la Máquina de Boltzmann Restringida (pre-RBM) muestran diferencias claras frente a la variable seguridad social. El clúster 0 agrupa principalmente a madres afiliadas al régimen contributivo. En contraste, el clúster 2 representa a una población con afiliación exclusiva al régimen subsidiado. Los clústeres 1 y 3 presentan mayor heterogeneidad, incluyendo afiliación al régimen subsidiado y posibles casos sin afiliación o con regímenes especiales.

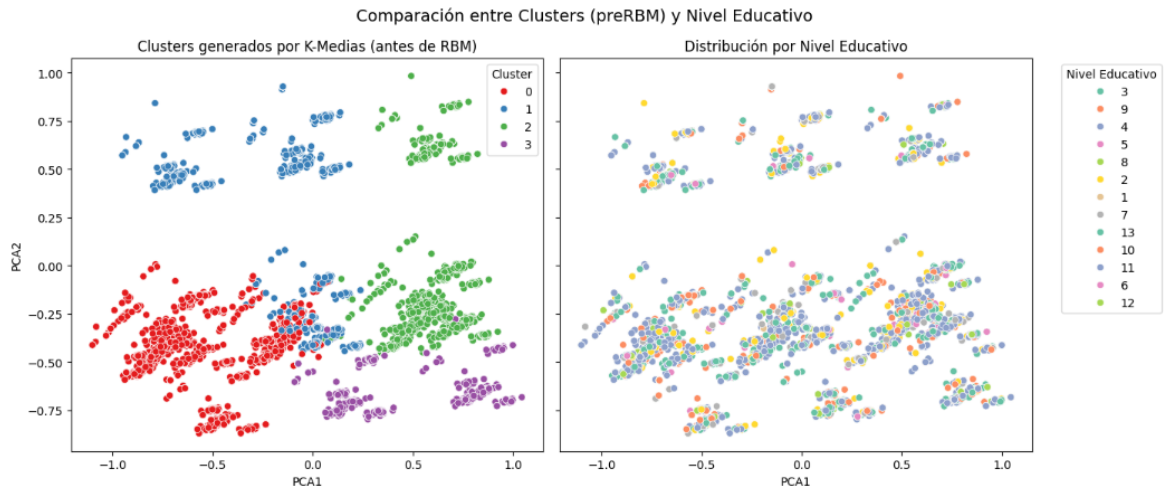


Figura 32. Distribución de clústeres K-means vs Nivel Educativo. Año 2021.

Como se observa en la Figura 32, los clústeres generados por K-medias antes de aplicar la RBM presentan diferencias claras frente al nivel educativo. El clúster 0 concentra individuos con niveles educativos más altos, mientras que los clústeres 2 y 3 agrupan predominantemente a personas con niveles más bajos. Por su parte, el clúster 1 muestra mayor diversidad, con presencia de varios niveles educativos, indicando heterogeneidad en este grupo.

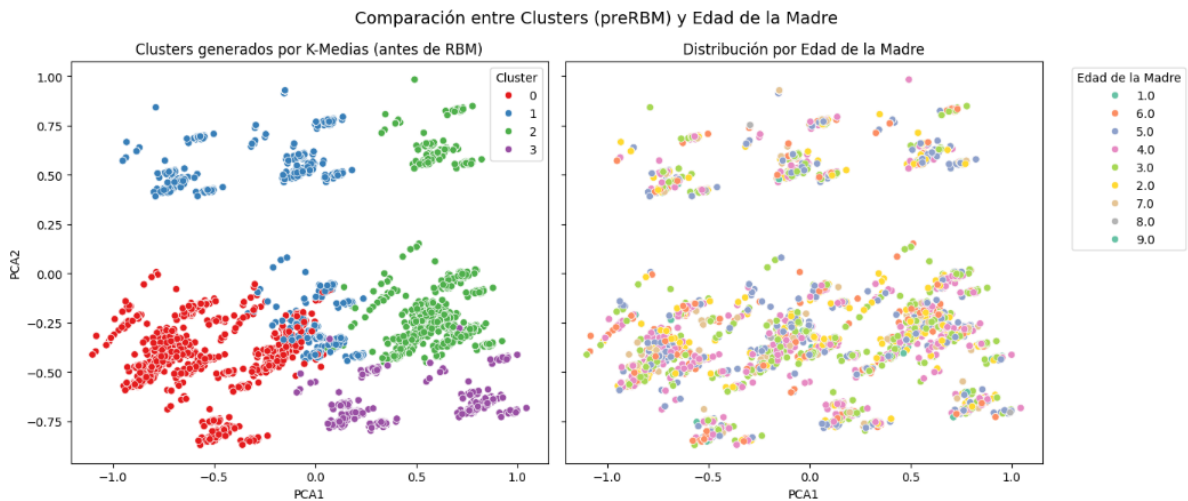


Figura 33. Distribución de clústeres K-means vs Edad Madre. Año 2021.

Como se muestra en la Figura 33, los clústeres generados antes de la RBM presentan una distribución relativamente homogénea respecto a la edad materna. Todos los clústeres contienen mayoritariamente madres entre los rangos 2 a 6, lo que indica que la edad no es una variable altamente discriminante entre los grupos. Sin embargo, se observan ligeras concentraciones de grupos más jóvenes en el clúster 0 y mayor presencia de madres mayores en el clúster 1.

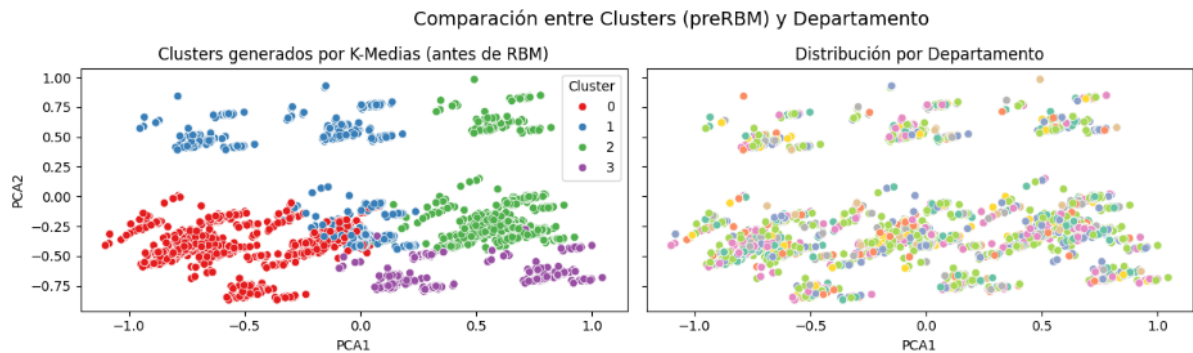


Figura 34. Distribución de clústeres K-means vs Código Departamento. Año 2021.

Como se observa en la figura 34, la variable departamento presenta una alta dispersión dentro de todos los clústeres generados por K-medias antes de la RBM. Esta distribución indica que los grupos contienen individuos de múltiples regiones del país, lo que sugiere que el lugar de procedencia no es un factor determinante en la segmentación inicial, aunque aporta diversidad territorial dentro de los clústeres.

Se realizó el cálculo de la entropía para evaluar la diversidad interna de cada agrupación según las variables.

Entropía por clúster respecto a 'cod_dpto' (antes de RBM):

Clúster 0: Entropía = 3.5551 bits

Clúster 1: Entropía = 3.4188 bits

Clúster 2: Entropía = 4.2537 bits

Clúster 3: Entropía = 4.3270 bits

Entropía por clúster respecto a 'edad_madre' (antes de RBM):

Clúster 0: Entropía = 2.4849 bits

Clúster 1: Entropía = 2.4587 bits

Clúster 2: Entropía = 2.5604 bits

Clúster 3: Entropía = 2.4625 bits

Entropía por clúster respecto a 'niv_edum' (antes de RBM):

Clúster 0: Entropía = 2.4503 bits

Clúster 1: Entropía = 0.5756 bits

Clúster 2: Entropía = 1.8208 bits

Clúster 3: Entropía = 0.0000 bits

Entropía por clúster respecto a 'seg_social' (antes de RBM):

Clúster 0: Entropía = 0.4438 bits

Clúster 1: Entropía = 1.0546 bits

Clúster 2: Entropía = 0.0179 bits

Clúster 3: Entropía = 0.7901 bits

A continuación, se presenta un gráfico de barras que resume la entropía calculada por clúster y por variable, permitiendo identificar visualmente los niveles de homogeneidad o heterogeneidad presentes en cada agrupación.

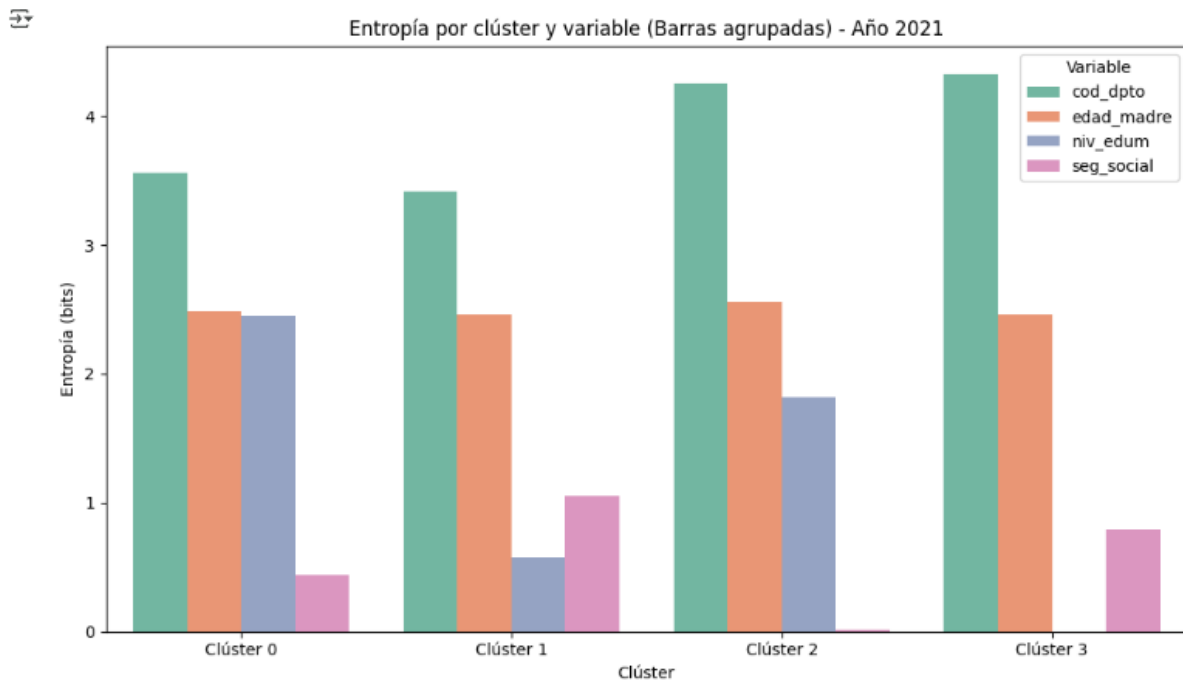


Figura 35. Entropía por clúster y variable socioeconómica. Año 2021.

Como se muestra en la Figura 35, los clústeres 0, 2 y 3 presentan altos niveles de entropía en la variable `cod_dpto`, lo que indica una gran diversidad territorial. El clúster 1, por su parte, evidencia menor diversidad en el nivel educativo, reflejando homogeneidad en esta característica. En cuanto a la seguridad social, se observa baja entropía en el clúster 2, lo que sugiere afiliaciones similares dentro del grupo. Estos patrones confirman la segmentación de los clústeres según condiciones sociales diferenciadas

2023

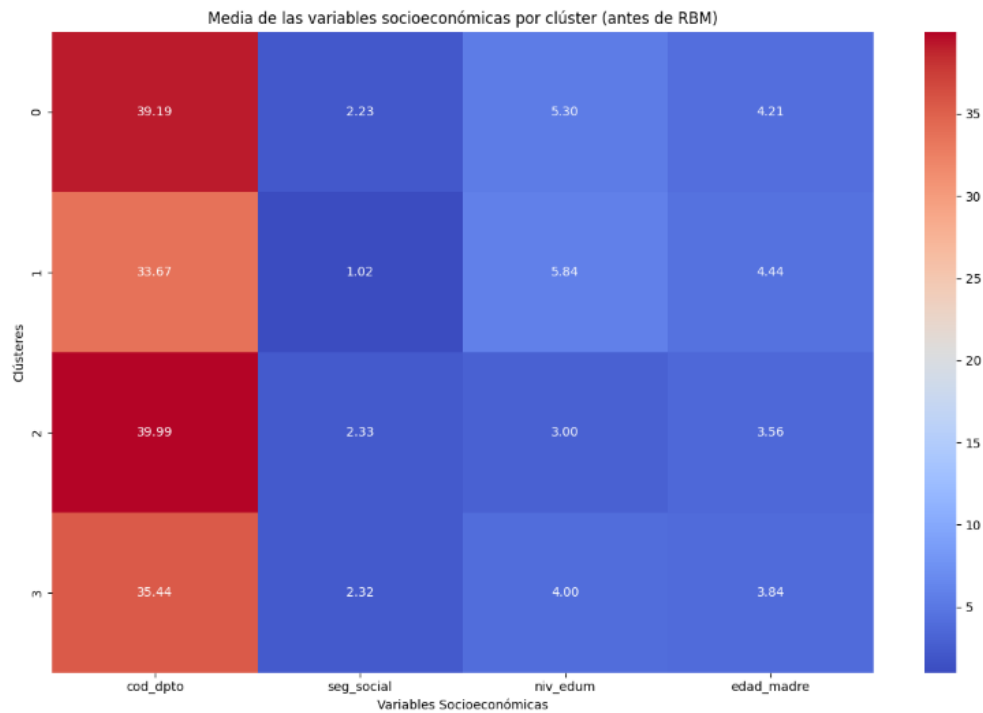


Figura 36. Agrupación por cluster para el año 2023.

Clúster 0:

- Edad de la madre: 4.21, correspondiente al grupo de 25 a 29 años.
- Nivel educativo: 5.30, indicando predominancia de madres con educación media técnica.
- Régimen de seguridad social: 2.23, lo que sugiere afiliación mayoritaria al régimen subsidiado.
- Departamento promedio: 39.19, asociado a zonas como Huila o La Guajira.

Este grupo representa madres jóvenes adultas con nivel medio educativo y cobertura subsidiada, aunque no es el grupo más vulnerable, presenta condiciones mixtas en cuanto a protección social.

Clúster 1:

- Edad de la madre: 4.44, lo que indica madres entre 25 y 29 años.
- Nivel educativo: 5.84, correspondiente a media técnica o normalista.
- Régimen de seguridad social: 1.02, lo que refleja afiliación predominante al régimen contributivo.
- Departamento promedio: 33.67, cercano a departamentos como Choco o Huila.

Este clúster refleja una población con mejores condiciones: mujeres adultas jóvenes, mayor nivel educativo y acceso al régimen contributivo. Se considera un grupo con menor vulnerabilidad relativa.

Clúster 2:

- Edad de la madre: 3.56, correspondiente al rango de 20 a 29 años.
- Nivel educativo: 3.00, lo que indica educación básica secundaria.
- Régimen de seguridad social: 2.33, lo que sugiere afiliación predominante al régimen subsidiado.
- Departamento promedio: 39.99, asociado a regiones con mayores desigualdades como La Guajira o Huila.

Este grupo representa a madres más jóvenes con bajo nivel educativo y baja protección social, lo que indica una mayor vulnerabilidad socioeconómica y posible riesgo obstétrico.

Clúster 3:

- Edad de la madre: 3.84, lo que indica edades entre 20 y 29 años, con tendencia a mayor juventud.

- Nivel educativo: 4.00, correspondiente a media académica o clásica.
- Régimen de seguridad social: 2.32, indicando afiliación principalmente al régimen subsidiado.
- Departamento promedio: 35.44, asociado a regiones como Choco o Huila.

Este clúster representa a mujeres jóvenes, con nivel educativo medio-bajo y afiliación subsidiada, lo cual refleja condiciones precarias y una población en situación de mayor riesgo.

Para complementar el análisis descriptivo de los clústeres generados por K-means, se presentan a continuación visualizaciones de su distribución en un espacio bidimensional. Estas gráficas permiten evaluar gráficamente la calidad de la agrupación y su relación con las variables socioeconómicas originales

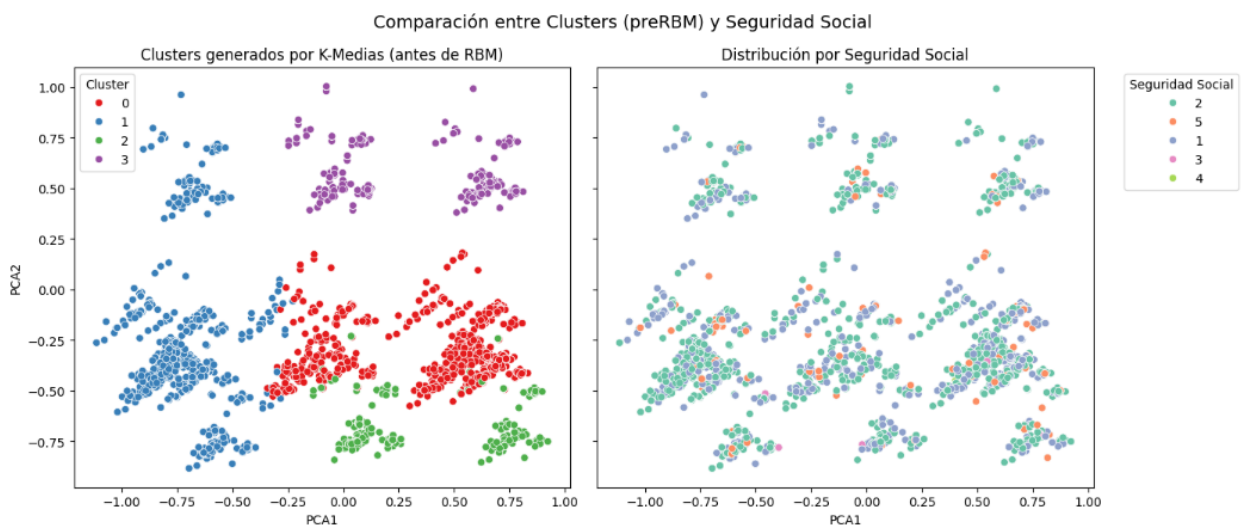


Figura 37. Distribución de clústeres K-means vs Seguridad Social. Año 2023.

La Figura 37 muestra que el clúster 0 agrupa en su mayoría a madres afiliadas al régimen contributivo. Por otra parte, el clúster 2 concentra una alta proporción de afiliadas al régimen subsidiado. Finalmente los clústeres 1 y 3 presentan mayor diversidad en los tipos de afiliación, combinando regímenes contributivo, subsidiado y especiales.

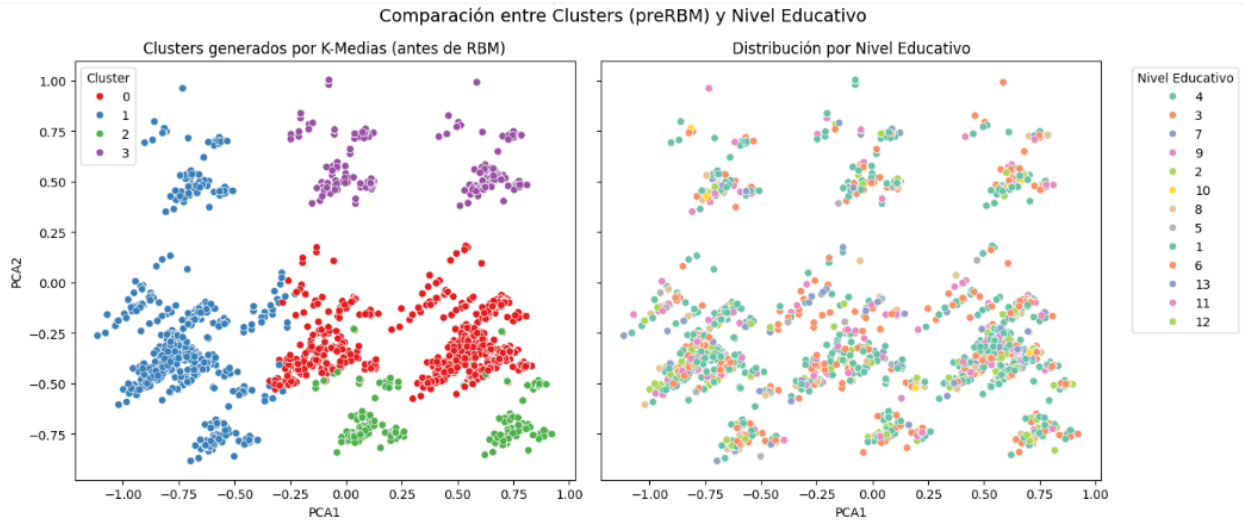


Figura 38. Distribución de clústeres K-means vs Nivel Educativo. Año 2023.

En la Figura 38 se puede observar que el clúster 0 presenta una distribución concentrada que corresponden a secundaria básica y media. El clúster 2 muestra predominancia de niveles bajos, asociados a educación básica primaria e incluso sin escolaridad completa. El clúster 1 refleja mayor diversidad de niveles educativos, con presencia tanto de categorías altas (universitaria) como bajas, lo que sugiere heterogeneidad en la escolaridad. Para finalizar el clúster 3 se asocia a niveles educativos más altos, correspondientes a formación técnica y superior.

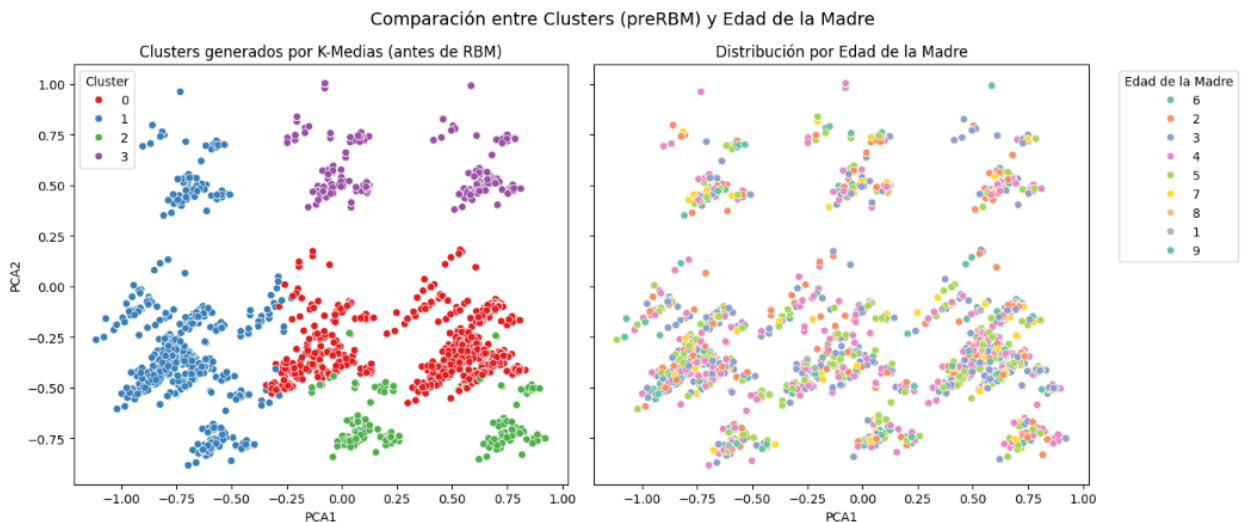


Figura 39. Distribución de clústeres K-means vs Edad Madre. Año 2023.

Como se puede observar en la Figura 39 el clúster 0 se concentra mayoritariamente en categorías 4 y 5, correspondientes a madres entre 20 y 29 años. El clúster 2 presenta un perfil similar, aunque con una mayor densidad en la categoría 4, que agrupa a madres de 20 a 24 años. El clúster 1 muestra mayor dispersión, con presencia de edades más tempranas (categoría 3) y algunas más avanzadas. Por otra parte, el clúster 3 evidencia una proporción relevante en las categorías 6 y 7, lo que indica mayor concentración de madres entre los 30 y 39 años.

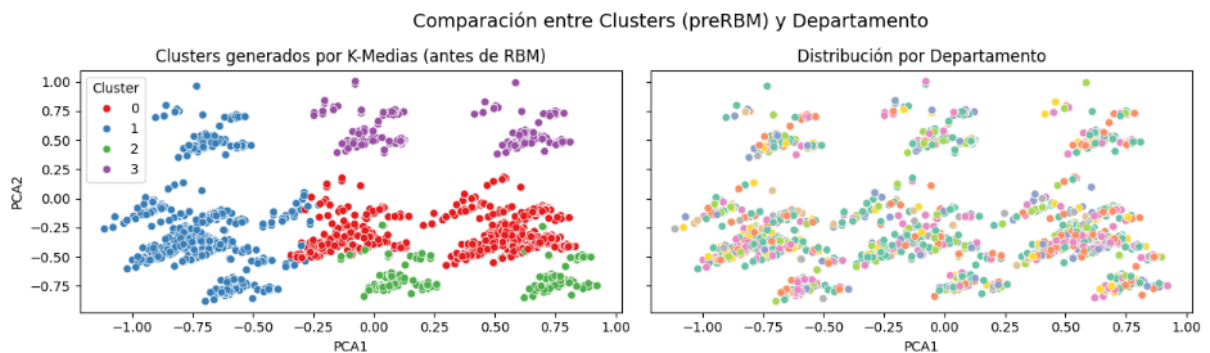


Figura 40. Distribución de clústeres K-means vs Código Departamento. Año 2023.

Como se puede observar en la Figura 40 el clúster 0 está distribuido principalmente en departamentos como Sucre, Bolívar y Córdoba. El clúster 1 presenta mayor dispersión, abarcando varios departamentos con diferente perfil territorial, lo que sugiere una mezcla heterogénea geográfica. El clúster 2 tiende a concentrarse en zonas específicas como Cauca y Nariño. Finalmente el clúster 3 aparece fuertemente agrupado en departamentos como Atlántico y Magdalena

Se realizó el cálculo de la entropía para evaluar la diversidad interna de cada agrupación según las variables.

Entropía por clúster respecto a 'cod_dpto' (antes de RBM):

Clúster 0: Entropía = 4.3557 bits

Clúster 1: Entropía = 3.3822 bits

Clúster 2: Entropía = 4.3068 bits

Clúster 3: Entropía = 3.8747 bits

Entropía por clúster respecto a 'edad_madre' (antes de RBM):

Clúster 0: Entropía = 2.6595 bits

Clúster 1: Entropía = 2.5106 bits

Clúster 2: Entropía = 2.5590 bits

Clúster 3: Entropía = 2.4039 bits

Entropía por clúster respecto a 'niv_edum' (antes de RBM):

Clúster 0: Entropía = 2.3219 bits

Clúster 1: Entropía = 2.5330 bits

Clúster 2: Entropía = 0.0000 bits

Clúster 3: Entropía = 0.0000 bits

Entropía por clúster respecto a 'seg_social' (antes de RBM):

Clúster 0: Entropía = 0.5241 bits

Clúster 1: Entropía = 0.0705 bits

Clúster 2: Entropía = 0.5451 bits

Clúster 3: Entropía = 0.5530 bits

A continuación, se presenta un gráfico de barras que resume la entropía calculada por clúster y por variable, permitiendo identificar visualmente los niveles de homogeneidad o heterogeneidad presentes en cada agrupación.

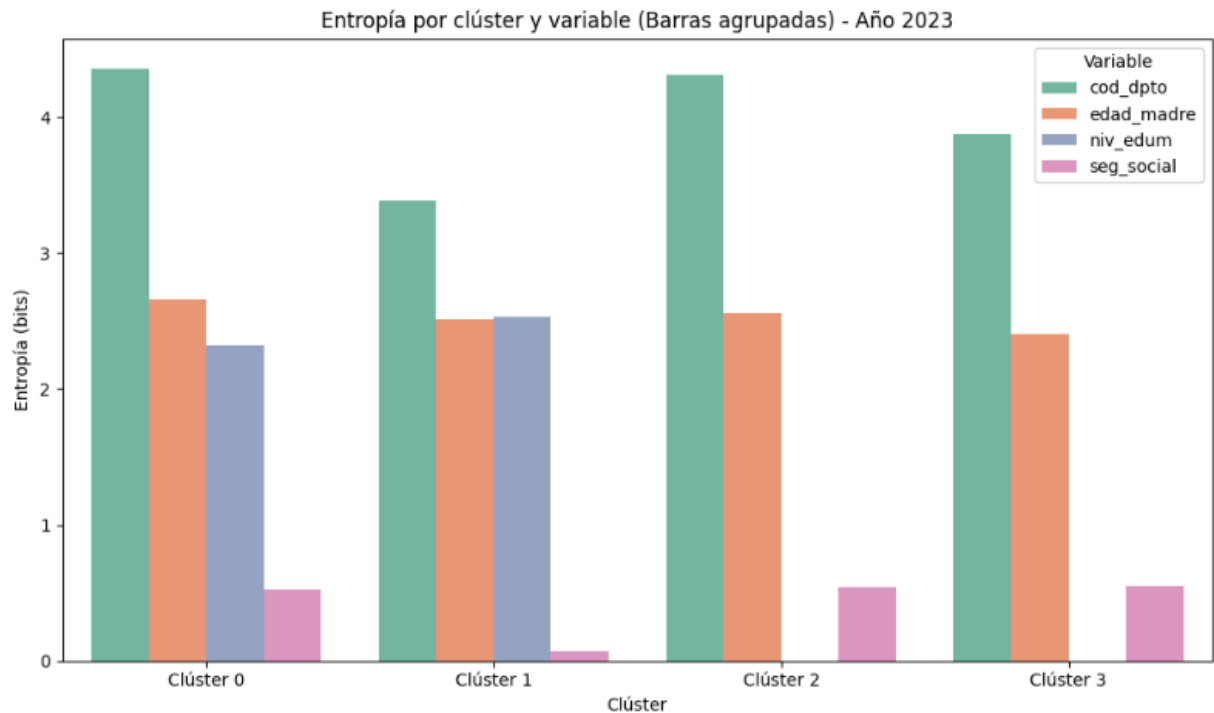


Figura 41. Entropía por clúster y variable socioeconómica. Año 2023.

La Figura 41 se puede observar que los clúster 2 y 3 presentan una entropía nula en la variable nivel educativo, reflejando afiliaciones homogéneas a un tipo específico de educación (Básica primaria o secundaria). Por otro lado, edad de la madre y seguridad social presentan menos diversidad en los clústeres 1 y 3. Finalmente, la variable código departamento presenta altos niveles de entropía en todos los clústeres, indicando una total dispersión territorial.

5.3.2. Después: K-Medias Además de la representación latente obtenida mediante la Máquina de Boltzmann Restringida, se aplicó el algoritmo de K-Medias directamente sobre estas representaciones, sin realizar previamente una reducción de dimensionalidad. El propósito fue evaluar si las características generadas por la RBM eran capaces de identificar agrupamientos significativos en los datos, permitiendo formar grupos con características socioeconómicas similares. Posteriormente, se utilizó el Análisis de Componentes Principales únicamente con fines de visualización, proyectando los clústeres en un espacio bidimensional para facilitar su

interpretación gráfica.

K-Medias luego de la RBM, permite analizar cómo las características ocultas que aprende la red neuronal pueden ayudar a descubrir patrones dentro de los datos, sin necesidad de contar con etiquetas o clasificaciones previas.

```
kmeans = KMeans(n_clusters=4, random_state=42)
kmeans.fit(datos_rbm) # Usar la variable correcta que contiene
las características transformadas por la RBM

# Asignar los clusters al DataFrame original
datos_importantes["Cluster"] = kmeans.labels_

# Ver las primeras filas con los clusters asignados
print(datos_importantes.head())

# Agrupar por cluster y calcular estadísticas
estadisticas_cluster = datos_importantes.groupby("Cluster").mean()

print("Estadísticas descriptivas por cluster:")
print(estadisticas_cluster)
```

Se usó el algoritmo K-Medias⁶⁴ para dividir los datos en cuatro grupos. Se eligió este número de grupos porque permite hacer una interpretación clara y útil de los resultados.

Después, se revisaron los promedios de las variables en cada grupo para entender sus diferencias. Con esto, se intentó identificar qué tipo de grupo social representa cada uno, según sus características.

⁶⁴ F. Pedregosa et al.: *Scikit-learn: Machine Learning in Python*. En: *Journal of Machine Learning Research* 12 (2011), págs. 2825-2830.

A continuación, se presenta y se analiza la importancia de las variables en cada clúster en los años 2020, 2021 y 2023, representada en un gráfico tipo *heatmap*.

Año 2020:

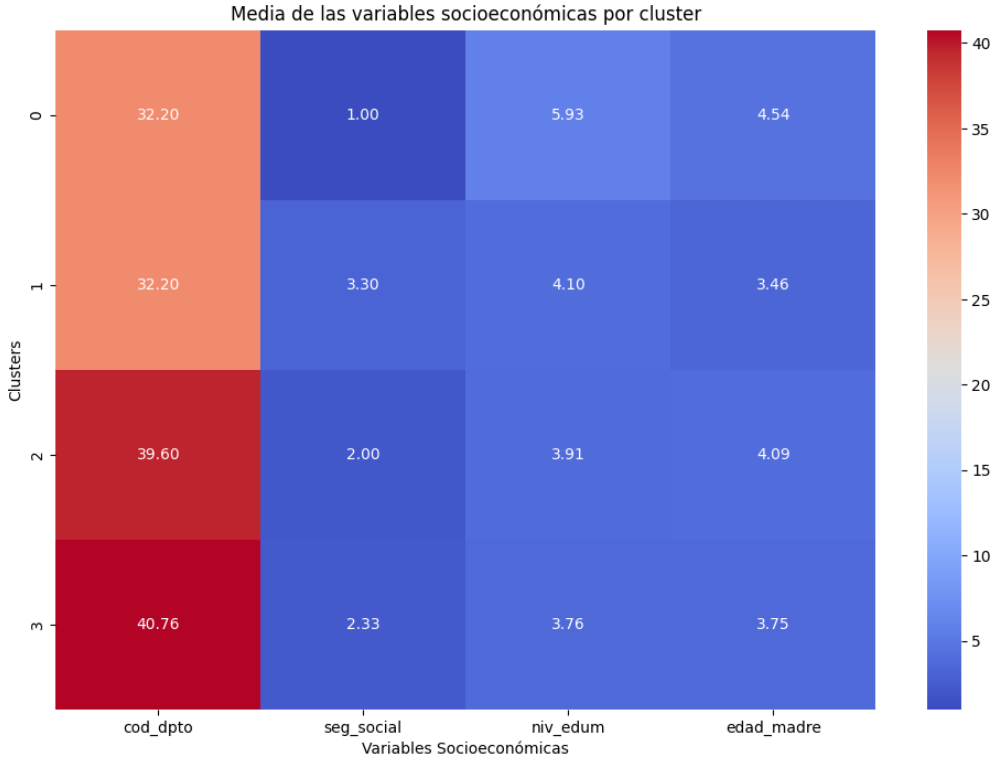


Figura 42. Agrupación por cluster para el año 2020.

Clúster 0:

- Edad de la madre: promedio de 4.54, correspondiente al grupo de 25 a 29 años.
- Nivel educativo: 5.93, lo cual indica una predominancia de madres con educación media técnica o normalista.
- Régimen de seguridad social: 1.00, afiliadas en su totalidad al régimen contributivo (EPS).
- Departamento promedio: 32.20, que corresponde a zonas como Cundinamarca.

Este grupo representa madres adultas jóvenes con educación media-alta y con sistema contributivo de salud. Al estar ubicadas en regiones con mayor desarrollo y mejores condiciones socioeconómicas, se considera una población con menor nivel de vulnerabilidad, donde las causas de muerte fetal podrían relacionarse con otros factores.

Clúster 1

- Edad de la madre: 3.46, lo que indica madres entre 20 y 24 años.
- Nivel educativo: 4.10, correspondiente a educación media académica.
- Régimen de seguridad social: 3.30, reflejando presencia de regímenes adaptado y especial.
- Departamento promedio: 32.20, similar al clúster anterior.

Aunque comparten zona geográfica con el grupo anterior, estas madres pertenecen a regímenes de seguridad social con altas tasas de mortalidad perinatal, como lo destaca el *Informe de Evento Mortalidad Perinatal y Neonatal Tardía: Colombia 2020* del Instituto Nacional de Salud⁶⁵. Esta población podría verse afectada por limitaciones en el acceso a servicios de salud, debido a las características particulares de los regímenes especiales a los que están afiliados.

Clúster 2

- Edad de la madre: 4.09, correspondiente a madres entre 25 y 29 años.
- Nivel educativo: 3.91, predominancia de madres con educación secundaria o básica completa.
- Régimen de seguridad social: 2.00, lo cual indica afiliación exclusiva al régimen subsidiado.
- Departamento promedio: 39.60, se aproxima a departamentos como Huila y la Guajira.

⁶⁵ Instituto Nacional de Salud: *Colombia. Mortalidad Perinatal y Neonatal Tardía*. En: *Instituto Nacional de Salud (2020)*. https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA_2020.pdf.

Este clúster refleja una población con condiciones sociales desfavorables, con bajo nivel educativo y afiliación al régimen subsidiado. Estas condiciones también están relacionadas con un mayor riesgo de muerte fetal, ya que factores como la pobreza, el bajo nivel educativo y las dificultades en el acceso a servicios de salud afectan la atención que reciben las madres durante el embarazo, lo que puede aumentar la probabilidad de que ocurran complicaciones graves.

Clúster 3

- Edad de la madre: 3.75, lo que corresponde a edades entre 20 y 29 años.
- Nivel educativo: 3.76, también dentro del rango de educación secundaria.
- Régimen de seguridad social: 2.33, indicando predominio del régimen subsidiado y posible presencia de no afiliación.
- Departamento promedio: 40.76, lo que sugiere cercanía a departamentos como Huila, La Guajira y Magdalena.

Este grupo tiene similitudes con el clúster 2, como un nivel educativo bajo, acceso limitado a servicios de salud y localización en regiones alejadas. Se trata de una población con condiciones sociales desfavorables, particularmente en áreas rurales o de difícil acceso.

Para complementar el análisis descriptivo de los clústeres generados por K-means, se presentan a continuación visualizaciones de su distribución en un espacio bidimensional. Estas gráficas permiten evaluar gráficamente la calidad de la agrupación y su relación con las variables socioeconómicas originales.

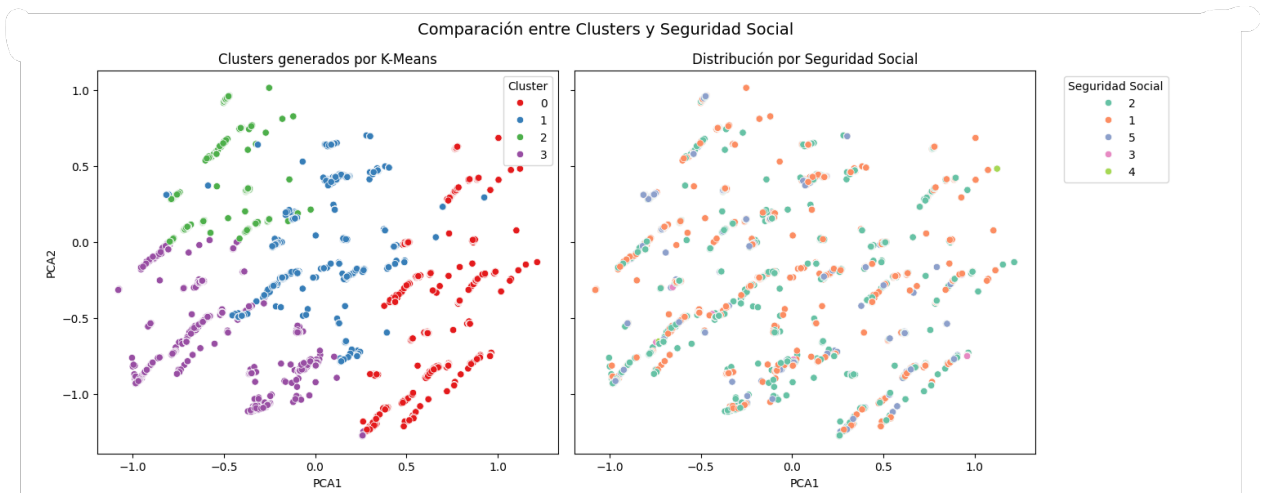


Figura 43. Distribución de clústeres K-means vs Seguridad Social. Año 2020.

Como se puede observar en la Figura 43, los clústeres generados por K-means presentan diferencias claras en cuanto al régimen de seguridad social, validando los patrones observados en el análisis numérico anterior. El clúster 0 muestra homogeneidad en la afiliación al régimen contributivo, mientras que el clúster 2 agrupa casos del régimen subsidiado.

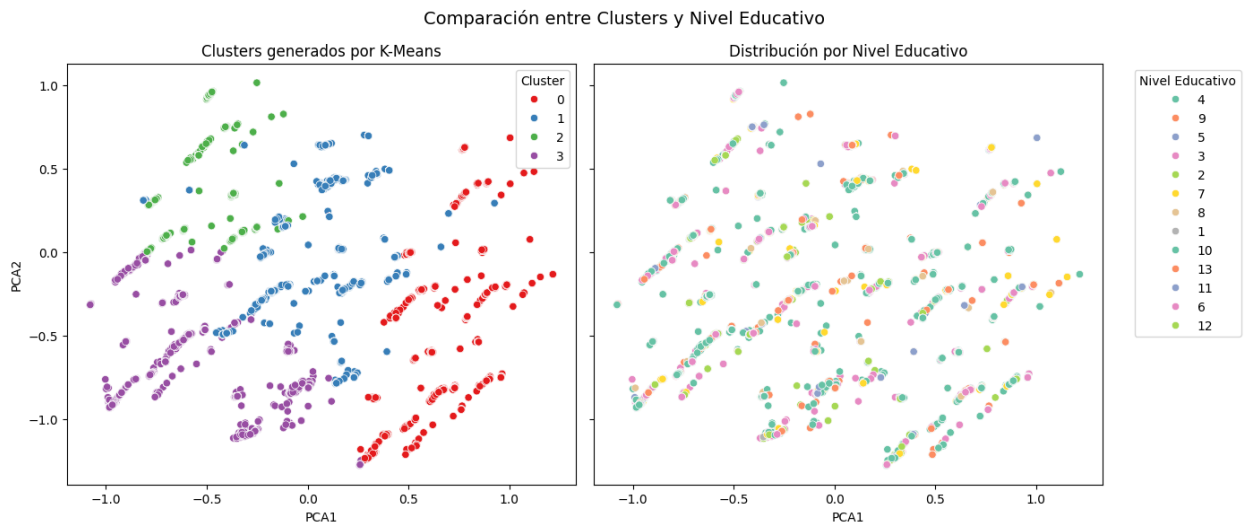


Figura 44. Distribución de clústeres K-means vs Nivel Educativo. Año 2020.

Como se puede observar en la Figura 44, los clústeres generados por K-means también muestran

diferencias significativas en relación con el nivel educativo de la madre. El clúster 0 concentra madres con educación media técnica o normalista, mientras que los clústeres 2 y 3 agrupan mayoritariamente a madres con niveles de educación básica o secundaria, consistentes con los resultados descriptivos.

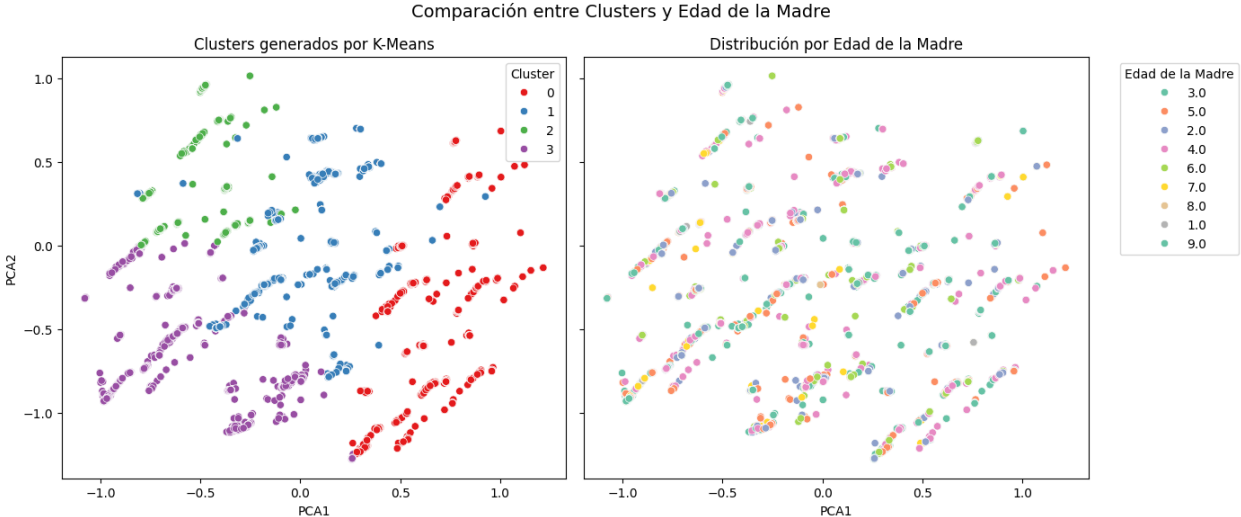


Figura 45. Distribución de clústeres K-means vs Edad de la Madre. Año 2020.

La Figura 45 evidencia que los clústeres presentan agrupaciones relacionadas con los rangos de edad de las madres. El clúster 0 se asocia a madres entre 25 y 29 años, mientras que el clúster 1 agrupa principalmente a madres más jóvenes, de entre 20 y 24 años. Esta distribución apoya las características detectadas en el análisis numérico previo.

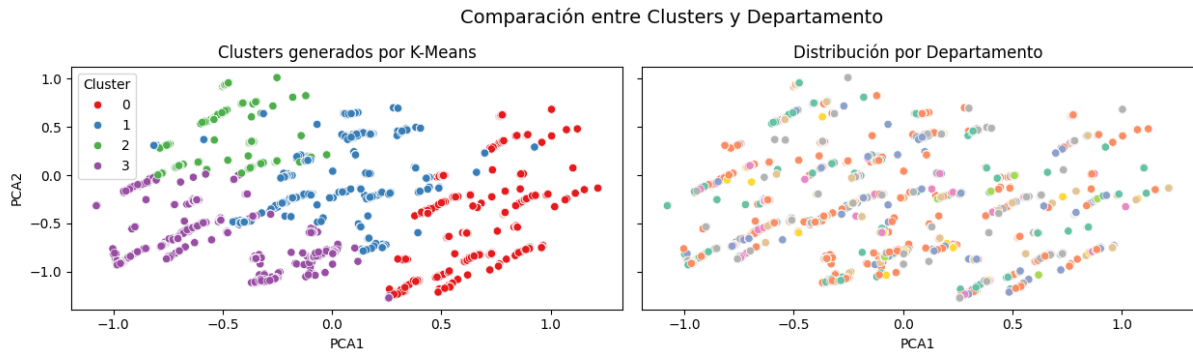


Figura 46. Distribución de clústeres K-means vs Departamento de Residencia. Año 2020.

Como se aprecia en la Figura 46, los clústeres formados por K-means presentan una distribución geográfica coherente con las interpretaciones anteriores. Los clústeres 2 y 3, asociados a condiciones socioeconómicas más desfavorables, se encuentran dispersos en departamentos históricamente más vulnerables, como Magdalena, Sucre, Huila y Nariño. En contraste, los clústeres 0 y 1 agrupan observaciones provenientes de departamentos con mejores indicadores de desarrollo, como Cundinamarca y Córdoba.

Para finalizar, como se pudo observar, los clústeres 2 y 3 reúnen a personas en situación más vulnerable, con bajo nivel educativo, afiliación al régimen subsidiado o sin acceso a salud, y que viven en zonas con mucha carencia de necesidades básicas.

El clúster 1 muestra una vulnerabilidad distinta, que esta relacionada con regímenes especiales de salud. Por su parte, el clúster 0 agrupa a personas con mejores condiciones, como mayor educación y afiliación al régimen contributivo, lo que puede estar relacionado con un menor riesgo, aunque aún se presentan los casos. Estos resultados señalan como principales factores de riesgo la baja escolaridad, el poco acceso a salud, la edad alta de la madre y las condiciones sociales difíciles.

Se realizó el cálculo de la entropía para evaluar la diversidad interna de cada agrupación según las variables.

Entropía por cluster respecto a 'cod_dpto':

Cluster 0: Entropía = 3.3662 bits

Cluster 1: Entropía = 3.6775 bits

Cluster 2: Entropía = 4.2052 bits

Cluster 3: Entropía = 4.4339 bits

Entropía por cluster respecto a 'edad_madre':

Cluster 0: Entropía = 2.3616 bits

Cluster 1: Entropía = 2.2077 bits

Cluster 2: Entropía = 1.9911 bits

Cluster 3: Entropía = 2.6111 bits

Entropía por cluster respecto a 'niv_edum':

Cluster 0: Entropía = 2.5662 bits

Cluster 1: Entropía = 1.2353 bits

Cluster 2: Entropía = 1.1337 bits

Cluster 3: Entropía = 2.1914 bits

Entropía por cluster respecto a 'seg_social':

Cluster 0: Entropía = 0.0000 bits

Cluster 1: Entropía = 1.6928 bits

Cluster 2: Entropía = 0.0000 bits

Cluster 3: Entropía = 0.6939 bits

A continuación, se presenta un gráfico de barras que resume la entropía calculada por clúster y por variable, permitiendo identificar visualmente los niveles de homogeneidad o heterogeneidad presentes en cada agrupación.

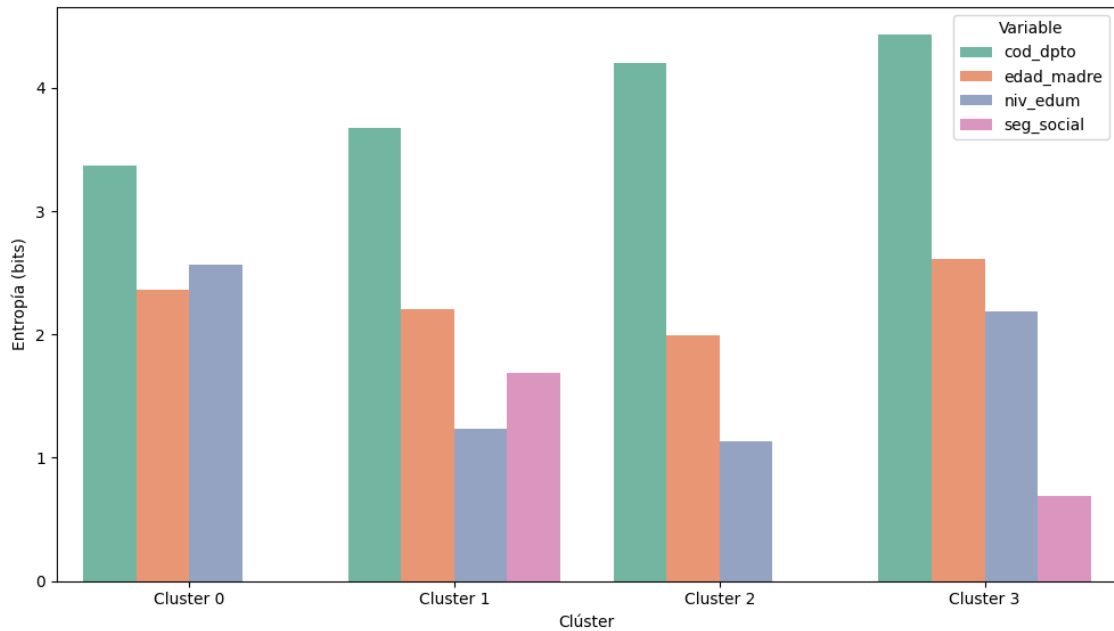


Figura 47. Entropía por clúster y variable socioeconómica. Año 2020.

Como se muestra en la Figura 47, los clústeres 0 y 2 presentan entropía nula en la variable de seguridad social, indicando afiliaciones homogéneas al régimen contributivo y subsidiado respectivamente. El clúster 1 muestra mayor diversidad en seguridad social y nivel educativo, mientras que el clúster 3 exhibe una mezcla más amplia de edades maternas y distribución geográfica. Estos resultados permiten entender mejor las diferencias internas en cada grupo y refuerzan su caracterización socioeconómica en relación con la mortalidad fetal.

Año 2021:

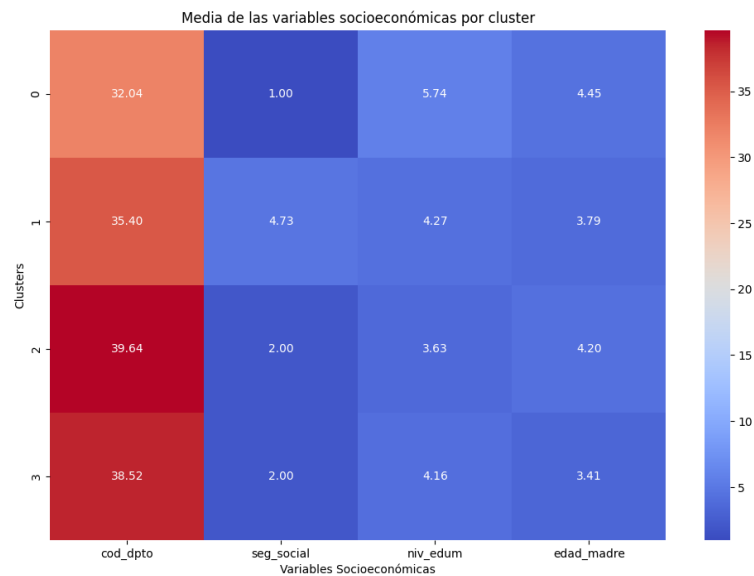


Figura 48. Agrupación por cluster para el año 2021.

Clúster 0:

- Edad de la madre: promedio de 4.45, correspondiente al grupo de 25 a 29 años.
- Nivel educativo: 5.74, lo que indica predominancia de madres con educación media técnica o profesional técnica.
- Régimen de seguridad social: 1.00, afiliación total al régimen contributivo.
- Departamento promedio: 32.04, cercano a zonas como Cundinamarca.

Este grupo representa a madres jóvenes con un nivel educativo medio-alto, afiliadas al sistema contributivo de salud y ubicadas en departamentos con condiciones sociales más favorables. Se considera una población con menor nivel de vulnerabilidad social, aunque igual con eventos de muerte fetal.

Clúster 1

- Edad de la madre: 3.78, lo que indica madres entre 20 y 24 años.
- Nivel educativo: 4.27, correspondiente a educación media académica.
- Régimen de seguridad social: 4.73, reflejando presencia importante de regímenes especiales o excepcionales.
- Departamento promedio: 35.40, influido principalmente por Cundinamarca.

Este grupo comparte algunas condiciones con el clúster 0 como el nivel educativo medio, pero se diferencia por la afiliación a regímenes especiales. Las particularidades de estos regímenes podrían implicar barreras en el acceso y la calidad de la atención en salud.

Clúster 2

- Edad de la madre: 4.20, lo que equivale a madres entre 25 y 29 años.
- Nivel educativo: 3.63, predominancia de educación básica secundaria.
- Régimen de seguridad social: 2.00, afiliación solo al régimen subsidiado.
- Departamento promedio: 39.64, se aproxima a departamentos como Huila y Magdalena.

Este clúster refleja una población con condiciones sociales desfavorables como bajo nivel educativo, afiliación al régimen subsidiado y ubicación en departamentos que han enfrentado carencias en el desarrollo social. Estas condiciones están relacionadas con un mayor riesgo de muerte fetal por las dificultades en el acceso a servicios de salud de calidad durante el embarazo.

Clúster 3

- Edad de la madre: 3.41, lo que corresponde a madres entre 20 y 24 años.
- Nivel educativo: 4.15, dentro del rango de educación media.

- Régimen de seguridad social: 2.00, afiliación única al régimen subsidiado.
- Departamento promedio: 38.52, cercano a departamentos como Huila y Magdalena.

Este grupo comparte características con el clúster 2, como la afiliación al régimen subsidiado y un nivel educativo medio-bajo.

Para complementar el análisis descriptivo de los clústeres generado para el año 2021, se presentan a continuación las visualizaciones que muestran la distribución de los grupos formados por K-means en relación con las variables socioeconómicas seleccionadas.

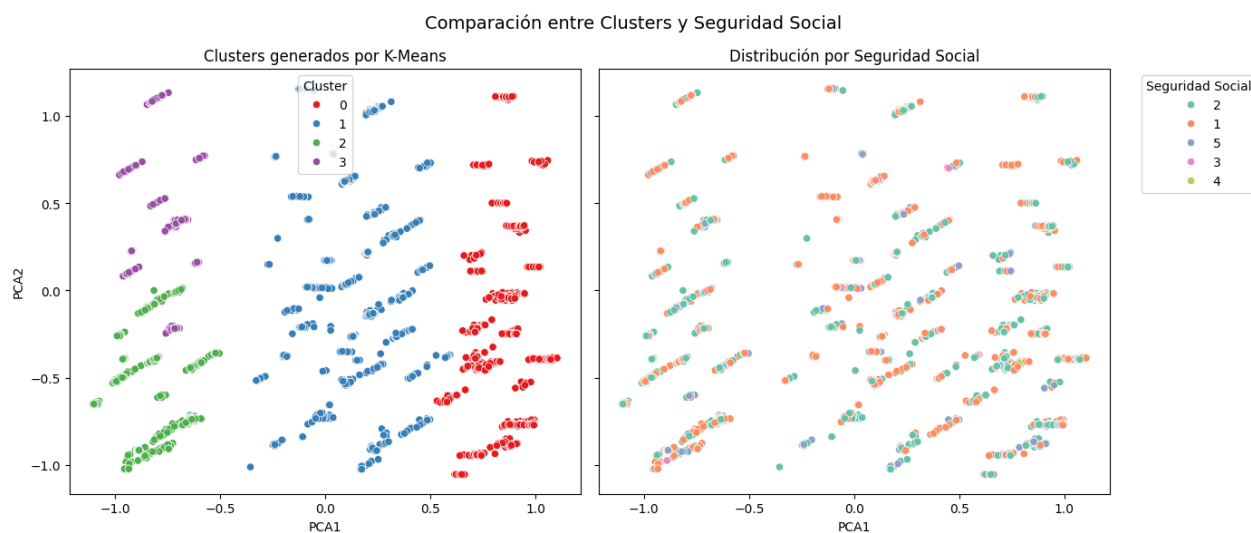


Figura 49. Distribución de clústeres K-means vs Seguridad Social. Año 2021.

Como se puede observar en la Figura 49, los clústeres generados por K-means muestran diferencias claras en cuanto al régimen de seguridad social. El clúster 0 concentra madres afiliadas al régimen contributivo (EPS), mientras que los clústeres 2 y 3 agrupan casos con afiliación al régimen subsidiado. El clúster 1 presenta una mayor heterogeneidad, asociada a la presencia de regímenes especiales o excepcionales.

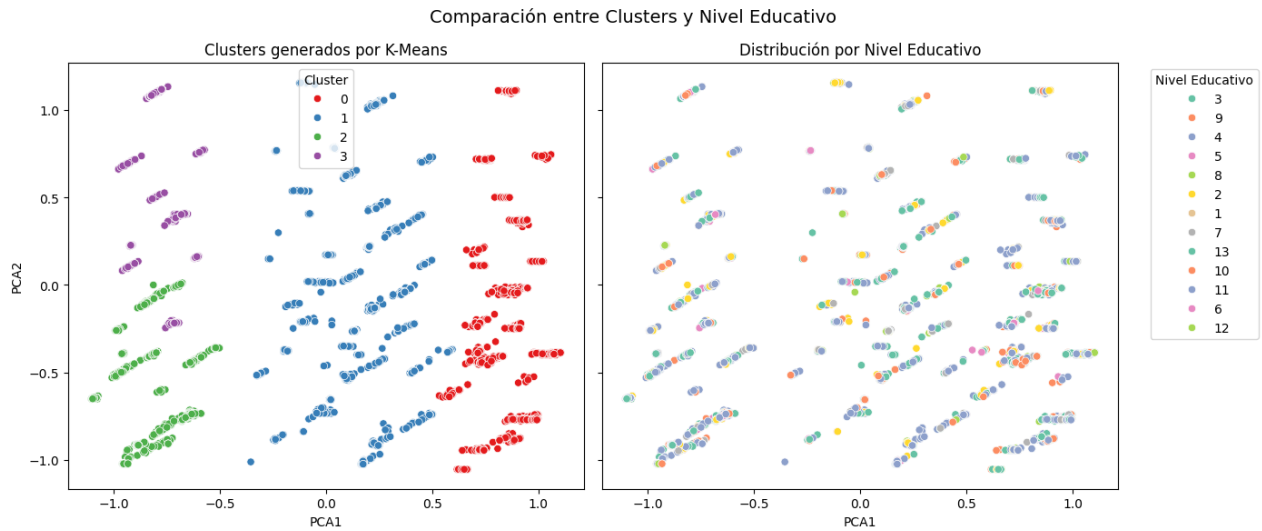


Figura 50. Distribución de clústeres K-means vs Nivel Educativo. Año 2021.

En la Figura 50 se evidencia que los clústeres difieren también en relación con el nivel educativo. El clúster 0 agrupa principalmente madres con educación media técnica o profesional, mientras que los clústeres 2 y 3 presentan niveles educativos más bajos, relacionados con educación básica secundaria. El clúster 1 refleja un perfil intermedio correspondiente a educación media académica.

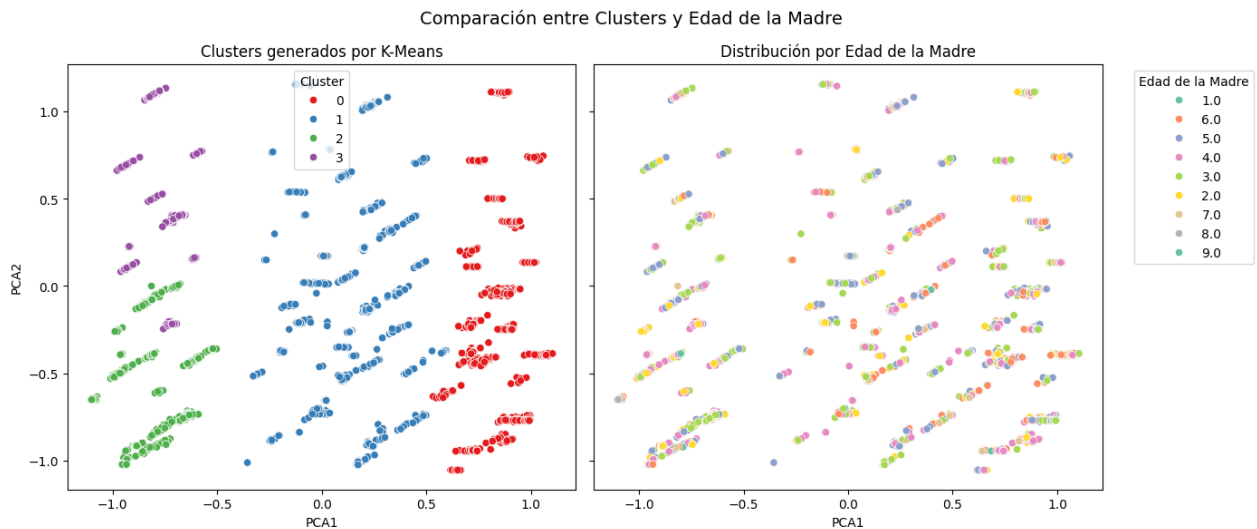


Figura 51. Distribución de clústeres K-means vs Edad de la Madre. Año 2021.

La Figura 51 muestra que el clúster 0 se asocia principalmente a madres entre 25 y 29 años, mientras que los clústeres 1 y 3 agrupan a madres más jóvenes, entre 20 y 24 años. Este comportamiento es coherente con las estadísticas descriptivas obtenidas previamente para el año 2021.

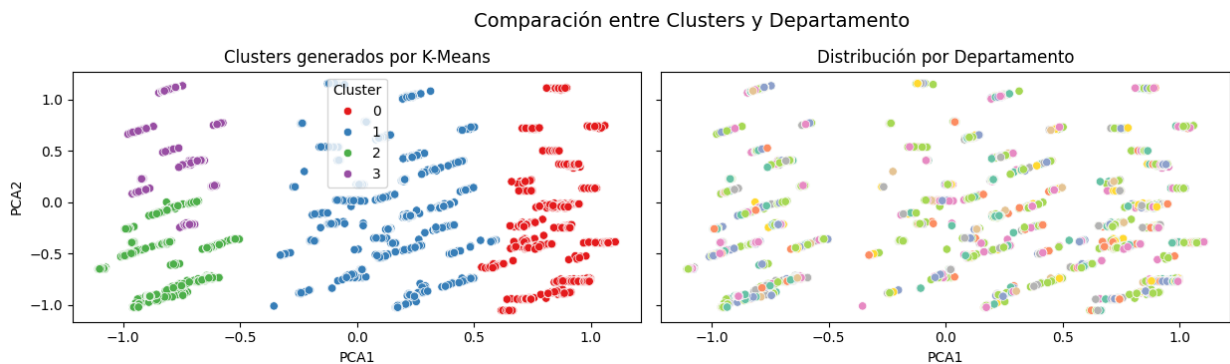


Figura 52. Distribución de clústeres K-means vs Departamento. Año 2021.

Finalmente, como se aprecia en la Figura 52, los clústeres presentan cierta dispersión territorial, aunque los clústeres 2 y 3 agrupan registros provenientes de departamentos asociados a mayores niveles de pobreza y necesidades básicas insatisfechas. El clúster 0, en contraste, se vincula a departamentos con condiciones sociales más favorables, como Cundinamarca o Boyacá.

En resumen, los clústeres 2 y 3 agrupan a poblaciones con condiciones sociales más desfavorables: bajo nivel educativo, afiliación al régimen subsidiado y ubicación en zonas con altos indicadores de pobreza. El clúster 1 representa una vulnerabilidad institucional asociada a regímenes especiales de salud. Finalmente, el clúster 0 refleja una población con mejores condiciones educativas y acceso al régimen contributivo.

Se realizó el cálculo de la entropía para evaluar la diversidad interna de cada agrupación según las variables.

Entropía por cluster respecto a 'cod_dpto':

Cluster 0: Entropía = 3.3830 bits

Cluster 1: Entropía = 3.8884 bits

Cluster 2: Entropía = 4.3215 bits

Cluster 3: Entropía = 4.2369 bits

Entropía por cluster respecto a 'edad_madre':

Cluster 0: Entropía = 2.4921 bits

Cluster 1: Entropía = 2.4722 bits

Cluster 2: Entropía = 2.2543 bits

Cluster 3: Entropía = 0.9768 bits

Entropía por cluster respecto a 'niv_edum':

Cluster 0: Entropía = 2.4880 bits

Cluster 1: Entropía = 2.2651 bits

Cluster 2: Entropía = 2.1492 bits

Cluster 3: Entropía = 2.0391 bits

Entropía por cluster respecto a 'seg_social':

Cluster 0: Entropía = 0.0000 bits

Cluster 1: Entropía = 0.5854 bits

Cluster 2: Entropía = 0.0000 bits

Cluster 3: Entropía = 0.0000 bits

A continuación, se presenta un gráfico de barras que resume la entropía calculada por clúster y por variable, permitiendo identificar visualmente los niveles de homogeneidad o heterogeneidad presentes en cada agrupación.

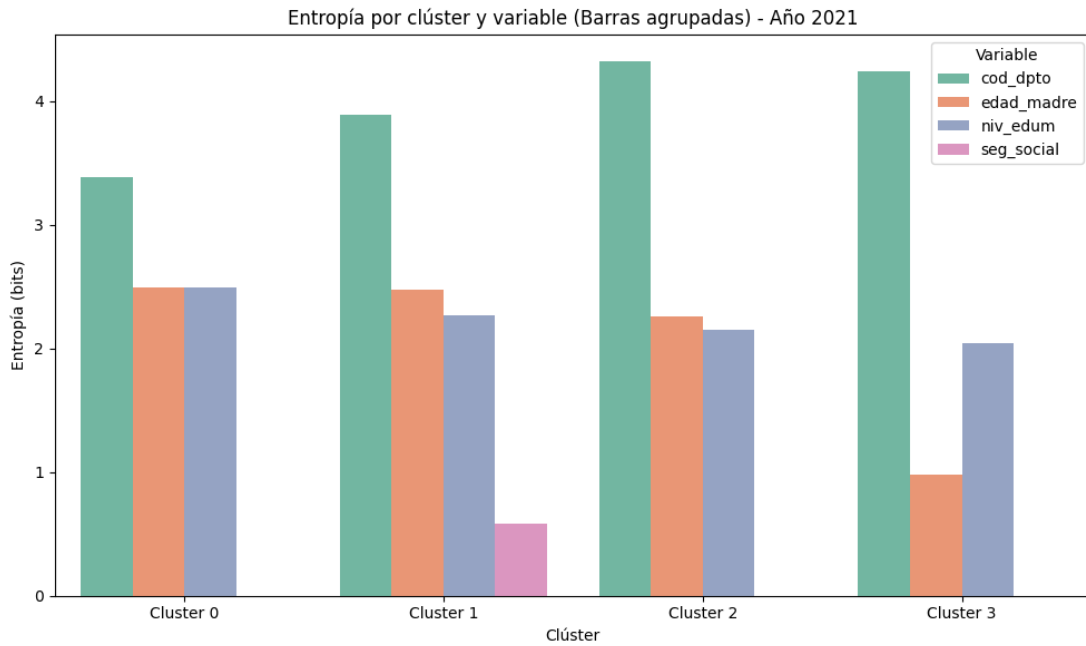


Figura 53. Entropía por clúster y variable socioeconómica. Año 2021.

La Figura 53 muestra que los clústeres 0, 2 y 3 presentan entropía nula en la variable seguridad social, reflejando afiliaciones homogéneas a un tipo específico de régimen (contributivo o subsidiado). En cuanto al nivel educativo y edad de la madre, se observa menor diversidad en los clústeres 2 y 3, mientras que el clúster 1 presenta una mayor dispersión educativa. La variable cod_dpto continúa mostrando altos niveles de entropía en todos los grupos, indicando dispersión territorial. Estos patrones respaldan la segmentación de las poblaciones en función de sus condiciones socioeconómicas.

Año 2023:

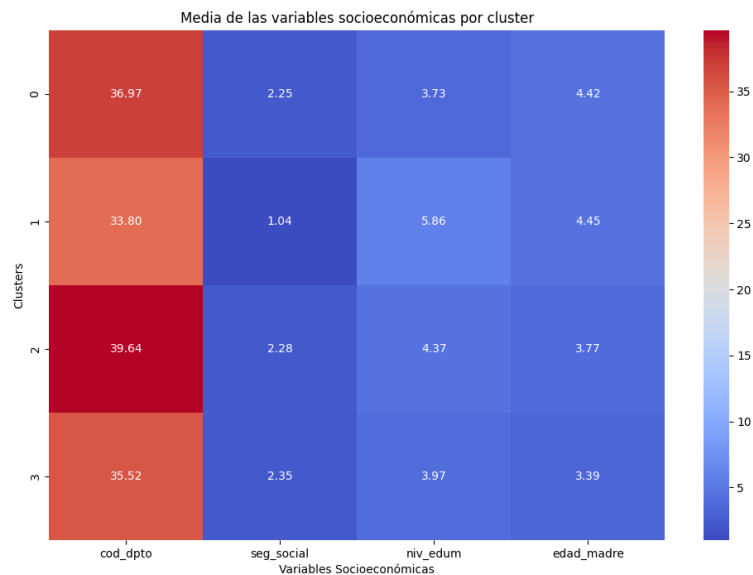


Figura 54. Agrupación por cluster para el año 2023.

Clúster 0:

- Edad de la madre: promedio de 4.42, correspondiente al grupo de 25 a 29 años.
- Nivel educativo: 3.73, destacan las madres con educación básica secundaria.
- Régimen de seguridad social: 2.25, lo que indica afiliación mayoritaria al régimen subsidiado, con posible presencia de no afiliación.
- Departamento promedio: 36.97, que aproxima a departamentos como Huila y La Guajira.

Este clúster agrupa a madres adultas pero jóvenes, con nivel educativo bajo y régimen de salud subsidiado. Están ubicadas en departamentos con un poco de desigualdad social y económica, lo que podría aumentar su riesgo de muerte fetal.

Clúster 1

- Edad de la madre: 4.45, correspondiente a edades entre 25 y 29 años.
- Nivel educativo: 5.86, lo cual indica un nivel medio técnico o superior no universitario.
- Régimen de seguridad social: 1.03, es decir, afiliación al régimen contributivo.
- Departamento promedio: 33.79, que corresponde a departamentos como Cundinamarca.

Este grupo representa una población con mejores condiciones: madres con mayor nivel educativo y afiliadas al régimen contributivo. Esto podría asociarse con menor vulnerabilidad socioeconómica, aunque los eventos de muerte fetal siguen presentes.

Clúster 2

- Edad de la madre: 3.76, correspondiente al grupo de 20 a 24 años.
- Nivel educativo: 4.36, que indica una mayor presencia de educación media académica.
- Régimen de seguridad social: 2.28, principalmente en el régimen subsidiado.
- Departamento promedio: 39.63, cercano a La Guajira o Magdalena.

Este grupo comparte características con el clúster 0, como nivel educativo medio-bajo y afiliación al régimen subsidiado. Están localizadas en regiones históricamente afectadas por desigualdad, lo que podría incrementar los factores de riesgo.

Clúster 3

- Edad de la madre: 3.39, lo que corresponde a edades entre 20 y 24 años.
- Nivel educativo: 3.97, dentro del rango de educación secundaria.
- Régimen de seguridad social: 2.35, indicando predominio del régimen subsidiado y posible no afiliación.

- Departamento promedio: 35.51, similar a zonas como El Choco.

Este grupo refleja condiciones sociales desfavorables, especialmente en contextos de zonas rurales, con menores oportunidades educativas y menor acceso a salud.

Para complementar el análisis descriptivo de los clústeres obtenidos en el año 2023, se presentan a continuación las visualizaciones que muestran la distribución de los grupos formados por K-means en relación con las variables socioeconómicas seleccionadas.

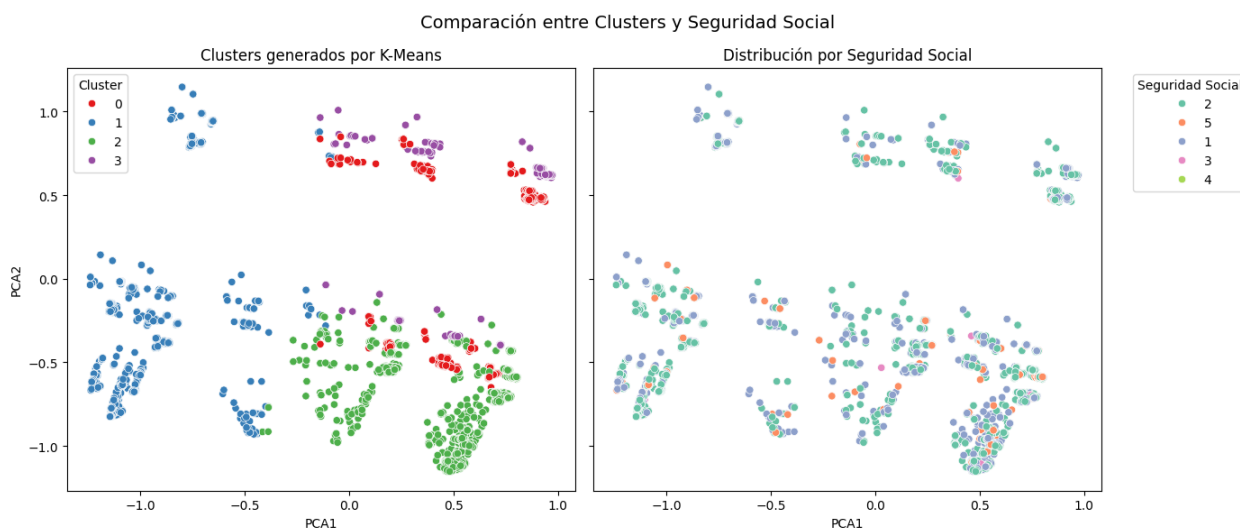


Figura 55. Distribución de clústeres K-means vs Seguridad Social. Año 2023.

La Figura 55 muestra que los clústeres 0, 2 y 3 presentan afiliaciones mayoritariamente al régimen subsidiado, mientras que el clúster 1 se agrupa claramente en torno al régimen contributivo, consistente con su perfil de mejores condiciones socioeconómicas.

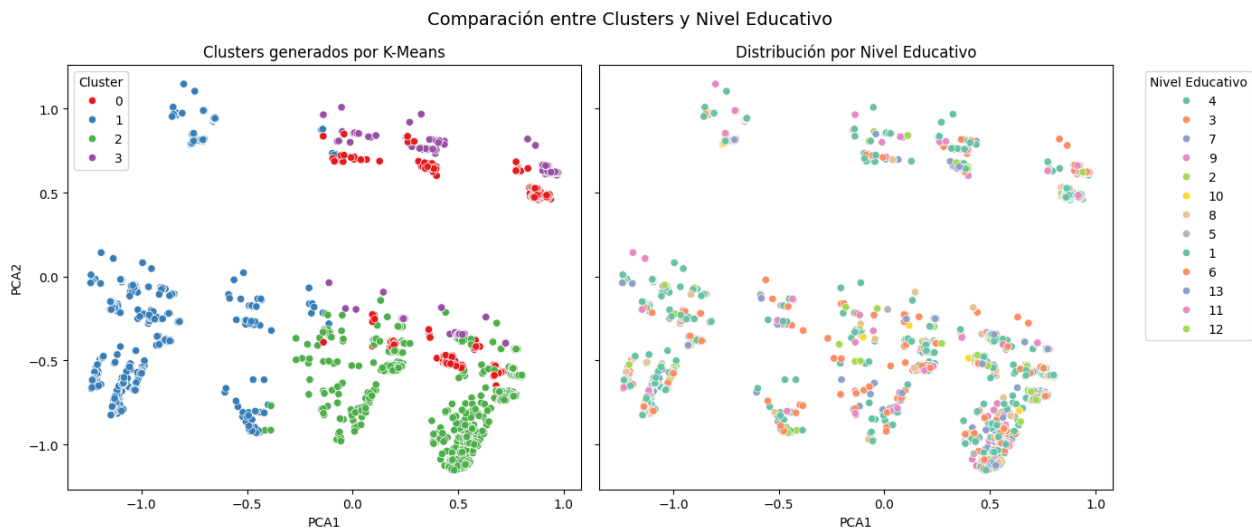


Figura 56. Distribución de clústeres K-means vs Nivel Educativo. Año 2023.

En la Figura 56 se observa que el clúster 1 agrupa madres con niveles de educación media técnica o profesional, mientras que los clústeres 0, 2 y 3 presentan niveles más bajos, mayormente asociados a educación secundaria o media académica.

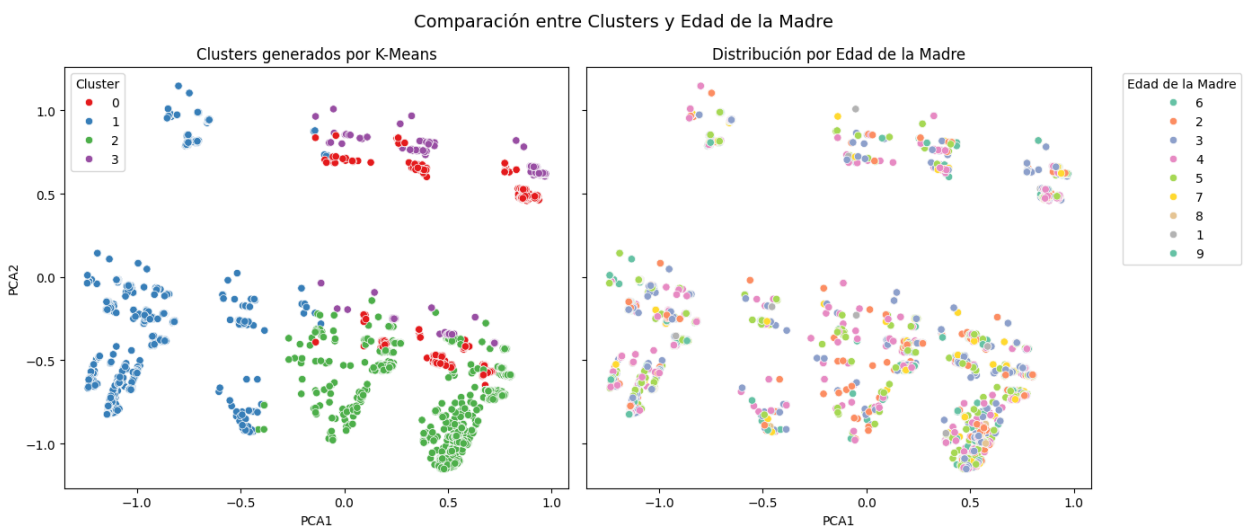


Figura 57. Distribución de clústeres K-means vs Edad de la Madre. Año 2023.

La Figura 57 evidencia que los clústeres 0 y 1 concentran madres de entre 25 y 29 años, mientras

que los clústeres 2 y 3 agrupan madres más jóvenes, de entre 20 y 24 años, lo cual coincide con el perfil descrito en el análisis de clústeres.

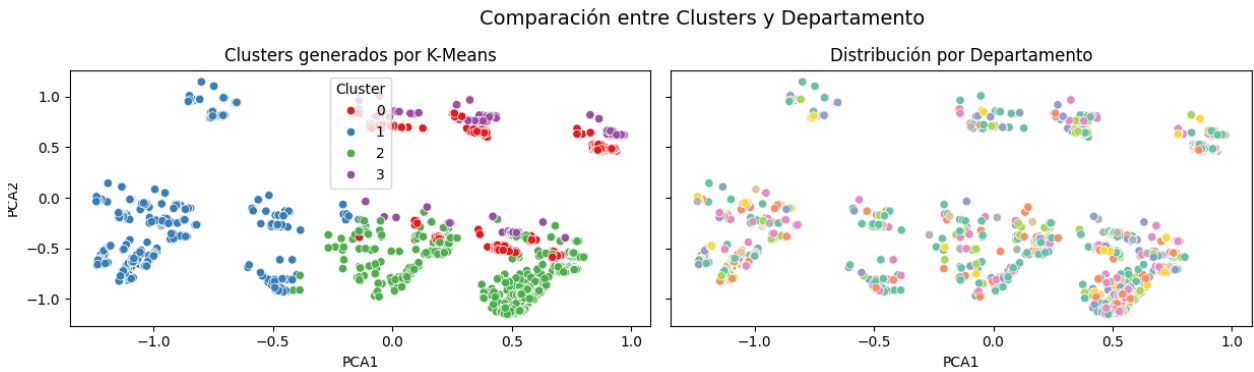


Figura 58. Distribución de clústeres K-means vs Departamento. Año 2023.

Finalmente, como se aprecia en la Figura 58, los clústeres 2 y 3 presentan una mayor dispersión geográfica, asociados a departamentos históricamente vulnerables como La Guajira, Magdalena y Atlántico, mientras que el clúster 1 se asocia a departamentos con mejores indicadores sociales.

Se realizó el cálculo de la entropía para evaluar la diversidad interna de cada agrupación según las variables.

Entropía por cluster respecto a 'cod_dpto':

Cluster 0: Entropía = 4.1212 bits

Cluster 1: Entropía = 3.4032 bits

Cluster 2: Entropía = 4.2845 bits

Cluster 3: Entropía = 3.9515 bits

Entropía por cluster respecto a 'edad_madre':

Cluster 0: Entropía = 2.0887 bits

Cluster 1: Entropía = 2.5130 bits

Cluster 2: Entropía = 2.5441 bits

Cluster 3: Entropía = 0.9660 bits

Entropía por cluster respecto a 'niv_edum':

Cluster 0: Entropía = 1.1642 bits

Cluster 1: Entropía = 2.5343 bits

Cluster 2: Entropía = 1.8372 bits

Cluster 3: Entropía = 0.6027 bits

Entropía por cluster respecto a 'seg_social':

Cluster 0: Entropía = 0.4717 bits

Cluster 1: Entropía = 0.1264 bits

Cluster 2: Entropía = 0.4855 bits

Cluster 3: Entropía = 0.5667 bits

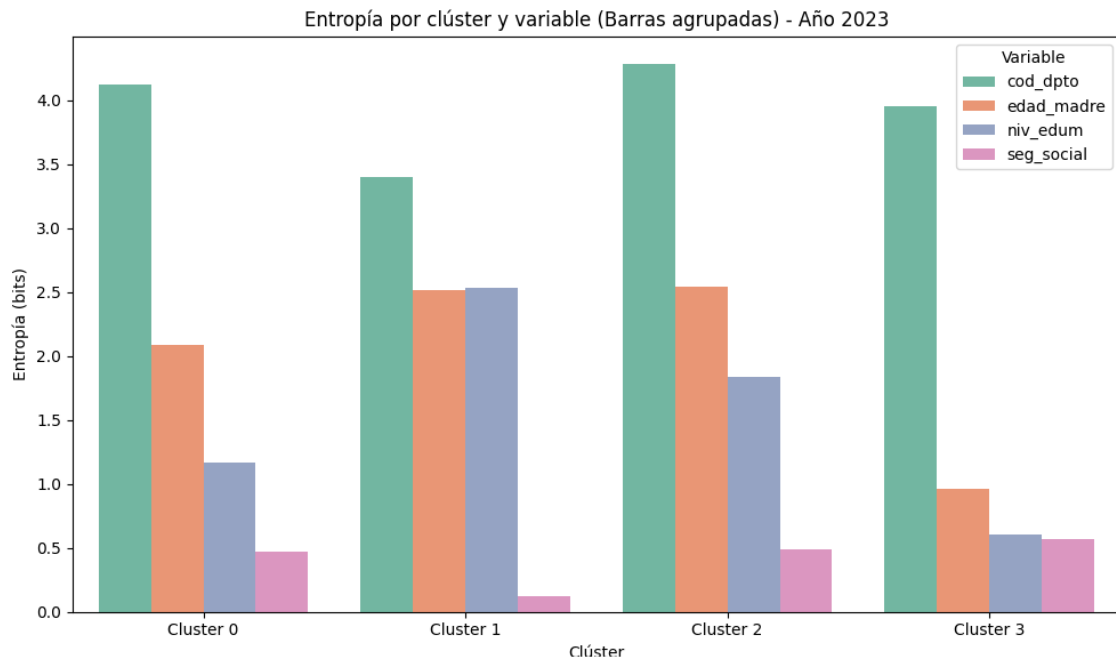


Figura 59. Entropía por clúster y variable socioeconómica. Año 2023.

La Figura 59 muestra que los clústeres 2 y 0 presentan mayores niveles de diversidad en las variables departamento y edad de la madre, indicando una mezcla más heterogénea de

condiciones sociales. El clúster 1 destaca por su homogeneidad en cobertura de seguridad social y en nivel educativo. El clúster 3, en contraste, muestra baja diversidad en edad y educación, indicando un grupo más uniforme en cuanto a estas características. Estos resultados refuerzan la segmentación de la población en grupos socioeconómicos diferenciados.

Concluyendo con el análisis e interpretación de K-Medias, en los tres años se reconocieron cuatro grupos que reflejan desigualdades socioeconómicas con implicaciones en las muertes fetales.

El primer grupo, o cluster 0, se caracterizó en todos los años por madres con mayor nivel educativo, afiliación al régimen contributivo y localización en departamentos con mejores condiciones sociales. Representa la población con menor vulnerabilidad, aunque no exenta de riesgo de muerte fetal.

El segundo grupo mostró un perfil intermedio en educación y edad, pero afiliado a regímenes especiales o adaptados, los cuales han sido asociados a mayores tasas de mortalidad en reportes oficiales⁶⁶. Este patrón fue evidente en 2020 y 2021, mientras que en 2023 no tanto, lo que podría sugerir que se hicieron cambios en la afiliación de salud.

Los “grupos más vulnerables”, o clústeres 2 y 3, se repitieron en los tres años, agrupando madres con bajo nivel educativo, afiliación al régimen subsidiado (o sin afiliación) y procedentes de departamentos con mayores carencias sociales. Estos clústeres concentran los mayores factores de riesgo para la muerte fetal, especialmente en zonas rurales.

En general, estos resultados sugieren que la mortalidad fetal en Colombia podría estar influenciada por factores sociales como el nivel educativo, el tipo de afiliación en salud y la región de residencia. Aunque se incluyó la edad materna como una de las variables de análisis, los resultados muestran que sus promedios por grupos se mantienen en rangos similares, especialmente entre los 20 y 29 años. Esto sugiere que la mayoría de los casos de muerte fetal

⁶⁶ Instituto Nacional de Salud: *Colombia. Mortalidad Perinatal y Neonatal Tardía*. En: *Instituto Nacional de Salud* (2020). https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA_2020.pdf.

se presentan en madres jóvenes o adultas tempranas. Por tanto la edad de la madre no parece ser un factor tan determinante para diferenciar perfiles de riesgo, en comparación con las otras variables. Pero, aunque la edad materna no mostró ser un factor discriminante en la agrupación, su interacción con condiciones socioeconómicas desfavorables podría incrementar los riesgos de muerte fetal.

6. CONCLUSIONES

Mediante la aplicación del algoritmo de K-Medias, dividido en dos etapas —**Antes** y **Después** de implementar la Máquina de Boltzmann Restringida (RBM)—, se observó que en los años seleccionados, al aplicar la RBM y obtener las variables latentes, el posterior uso del algoritmo de K-Medias permitió identificar clústeres mejor separados y definidos. Esto favoreció una visualización más clara y estructurada de patrones, facilitando la identificación de las variables socioeconómicas más influyentes, así como las categorías específicas que más inciden en las muertes fetales en Colombia. Adicionalmente, se evaluó la entropía global en ambas etapas, y tras la aplicación de la RBM se registraron, niveles de entropía más bajos y homogéneos entre los clústeres, lo que indica una mayor consistencia y calidad en el agrupamiento.

En resumen, la implementación de la RBM previa al clustering generó una mejora significativa, permitió reducir la dimensionalidad, extraer representaciones latentes que resaltan la relevancia de cada factor socioeconómico en el riesgo de muerte fetal, y mejorar las métricas de entropía, que presentaron valores inferiores en comparación con la etapa **Antes**, lo cual es un indicador positivo de desempeño. Posteriormente se realizó un análisis del antes y después de las representaciones generadas por la Máquina de Boltzmann Restringida, la cual permitió encontrar diferentes patrones socioeconómicos en las madres que tuvieron eventos de muerte fetal en Colombia durante los años 2020, 2021 y 2023.

En cada año analizado se identificaron de manera consistente cuatro grupos, que reflejan importantes desigualdades sociales. El **clúster 0**, en todos los periodos de la etapa **Después**, agrupó a madres con mayor nivel educativo, afiliadas al régimen contributivo de salud y que residen en departamentos con mejores condiciones de vida. Estas madres representan una población menos vulnerable, aunque no están completamente exentas del riesgo de muerte fetal. Al comparar con la etapa **Antes** de aplicar la RBM, se conformaron los 4 grupos, pero el **clúster 0** mostraba una composición más heterogénea y menos definida, con mayor mezcla de características socioeconómicas. La representación de madres con alto nivel educativo y afiliación

al régimen contributivo no era tan clara, lo que indica que la agrupación no capturaba con precisión los patrones latentes presentes en los datos.

El **clúster 1** mostró un perfil intermedio: madres con educación media y afiliadas a regímenes especiales o adaptados de salud, los cuales, según informes oficiales, están relacionados con mayores tasas de mortalidad perinatal. Este patrón fue especialmente evidente en 2020 y 2021, años marcados por la emergencia sanitaria de la pandemia de COVID-19, donde el acceso desigual a los servicios de salud se vio aún más acentuado. En 2023, se observó una disminución en la presencia de estos regímenes especiales, lo que podría reflejar esfuerzos en mejorar la cobertura en salud o cambios estructurales en los sistemas de afiliación, posiblemente como respuesta a las deficiencias evidenciadas durante la pandemia. Al observar con la etapa **Antes**, se nota que la caracterización del **clúster 1** era más difusa. Las madres afiliadas a regímenes especiales no se agrupaban de forma tan nítida, y su relación con el nivel educativo medio no era tan evidente. La falta de una representación clara limitaba la posibilidad de establecer asociaciones sólidas entre estas condiciones y el riesgo de muerte fetal.

Por su parte, los **clústeres 2 y 3** agruparon de forma repetida a las madres más vulnerables: aquellas con bajo nivel educativo, afiliadas al régimen subsidiado o sin afiliación, y que residen en departamentos históricamente afectados por pobreza y desigualdad. La vulnerabilidad de estos grupos pudo haberse intensificado durante los años de pandemia, debido a las restricciones en el acceso a servicios médicos y al deterioro de las condiciones socioeconómicas en zonas rurales, factores que incrementan el riesgo de eventos como la muerte fetal a comparación de la etapa **Antes** de aplicar la RBM, estos perfiles de alta vulnerabilidad no se distinguían con la misma claridad. Las madres en situación crítica aparecían dispersas entre distintos clústeres, dificultando la identificación de patrones consistentes. Además, la mezcla de características socioeconómicas dentro de cada grupo limitaba la posibilidad de focalizar intervenciones.

Aunque se analizó la edad de las madres, el promedio de edad en todos los grupos fue similar (entre 20 y 29 años), lo que sugiere que, por sí sola, la edad no fue un factor clave para diferenciar los perfiles de riesgo. Sin embargo, cuando la juventud materna se combina con condiciones

sociales desfavorables (como bajo nivel educativo y falta de acceso efectivo a la salud), puede aumentar la vulnerabilidad durante el embarazo.

Estos resultados confirman que factores como la educación, el tipo de afiliación en salud y la región de residencia son determinantes en el riesgo de muerte fetal en Colombia. Además, destacan cómo los efectos de la pandemia de COVID-19 pudieron influir en agravar las desigualdades sociales preexistentes, subrayando la necesidad urgente de implementar políticas públicas orientadas a cerrar las brechas en salud y garantizar un acceso equitativo a los servicios de salud en todo el país.

BIBLIOGRAFÍA

Abdi, Hervé y Lynne J. Williams: *Principal Component Analysis*. En: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.4 (2010), págs. 433-459. <https://personal.utdallas.edu/~herve/abdi-awPCA2010.pdf>.

Abdi, Hervé y Lynne J. Williams: *Principal component analysis*. En: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.4 (2010), págs. 433-459. DOI: 10.1002/wics.101.

Álvarez, María Alejandra et al.: *Diseño de un aplicativo para la identificación del estado de riesgo biopsicosocial en madres gestantes*. En: *Pontificia Universidad Javeriana* (2019). Trabajo de grado en modalidad de aplicación, Ingeniería Industrial, Primer Semestre 2019.

Arrieta Rodríguez, Eugenia Luz: *Predicción Temprana de Morbilidad Materna Extrema Usando Aprendizaje Automático*. En: *Universidad Tecnológica de Bolívar* (2017). Tesis de Maestría.

Bousquet, O., U. von Luxburg y G. Rätsch: *Advanced Lectures on Machine Learning*. Springer, 2003.

El-Brawany, Mohamed A. et al.: *Artificial intelligence-based data-driven prognostics in industry: A survey*. En: *Elsevier* 184 (2023).

Caicedo, E. F. y J. A. López: *Una aproximación práctica a las redes neuronales artificiales*. En: *Programa Editorial Universidad del Valle* (2017).

Cantero Lorenzo, J.: *Máquina de aprendizaje y aplicaciones*. En: *Universitat de les Illes Balears* (2022).

Clevert, Djork-Arne, Thomas Unterthiner y Sepp Hochreiter: *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. En: *International Conference on Learning Representations (ICLR)* (2016).

Colombia. *Mortalidad Perinatal y Neonatal Tardía*. En: *Instituto Nacional de Salud* (2020). https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA_2020.pdf.

Colombia. *Mortalidad Perinatal y Neonatal Tardía*. En: *Instituto Nacional de Salud* (2021). <https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA%20INFORME%202021.pdf>.

DANE: *Estadísticas vitales: Defunciones fetales 2020*. En: *Microdatos DANE* (2020). <https://microdatos.dane.gov.co/index.php/catalog/775>.

DANE: *Estadísticas vitales: Defunciones fetales 2021*. En: *Microdatos DANE* (2021). <https://microdatos.dane.gov.co/index.php/catalog/812>.

DANE: *Estadísticas vitales: Defunciones fetales 2023*. En: *Microdatos DANE* (2023). <https://microdatos.dane.gov.co/index.php/catalog/843>.

Dane, estadísticas vitales: *Defunciones fetales y no fetales. 3er trimestre 2023pr, año corrido 2023pr, cifras definitivas 2022*. En: *DANE* (2023). <https://www.dane.gov.co/files/operaciones/EEVV/bol-EEVV-Defunciones-IIItrim2023.pdf>.

Explicación de las funciones de pérdida en el machine learning. En: *Datacamp* (2024). <https://www.datacamp.com/es/tutorial/loss-function-in-machine-learning>.

Fuente Crespo, Laura de la: *Análisis de Componentes Principales*. En: *Recurso en línea* (s.f.). https://www.fuenterrebollo.com/Master-Econometria/Componentes_Principales.pdf.

Gregory, E. C., C. P. Valenzuela y D. L. Hoyert: *Fetal mortality: United States*. En: *National Vital Statistics Reports* (2020).

Hastie, T., R. Tibshirani y J. Friedman: *The Elements of Statistical Learning*. Springer, 2017.

Haykin, Simon: *Neural Networks and Learning Machines*. 3rd. Pearson Education, 2009.

Hinton, Geoffrey: *A Practical Guide to Training Restricted Boltzmann Machines*. En: *Department of Computer Science, University of Toronto* (2010).

IBM: *¿Qué es una función de pérdida?* En: *IBM Think* (2023). <https://www.ibm.com/es-es/think/topics/loss-function>.

Instituto Nacional de Salud: *Colombia. Mortalidad Perinatal y Neonatal Tardía*. En: *Instituto Nacional de Salud* (2020). https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA_2020.pdf.

Instituto Nacional de Salud: *Colombia. Mortalidad Perinatal y Neonatal Tardía*. En: *Instituto Nacional de Salud* (2021). <https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA%20INFORME%202021.pdf>.

Instituto Nacional de Salud (INS): *Informe de Evento: Mortalidad Perinatal y Neonatal Tardía 2020*. En: *INS* (2021). https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA_2020.pdf.

Jain, A. K.: *Data clustering: 50 years beyond K-means*. En: *Pattern Recognition Letters* 31.8 (2010), págs. 651-666.

- Jain, A. K. y R. C. Dubes: *Algorithms for Clustering Data*. Prentice-Hall, 1988.
- Jolliffe, Ian T.: *Principal Component Analysis*. 2nd. New York: Springer, 2002.
- Larrañaga, Pedro, Iñaki Inza y Abdelmalik Moujahid: *Redes Neuronales*. En: *Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad del País Vasco–Euskal Herriko Unibertsitatea* (s.f.).
- Lopez, Giulio et al.: *Entropy, Free Energy, and Work of Restricted Boltzmann Machines*. En: *Entropy* 22.5 (2020), pág. 538. DOI: 10.3390/e22050538. <https://www.mdpi.com/1099-4300/22/5/538>.
- MacQueen, J.: *Some methods for classification and analysis of multivariate observations*. En: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1 (1967), págs. 281-297.
- Máquina de Boltzmann restringida*. En: *Recurso en línea* (2023). <https://www.alexanderthamm.com/es/data-science-glossary/maquina-de-boltzmann-restringida/>.
- Maquina de Soporte Vectorial (SVM)*. En: *Recurso en línea* (2020). <https://medium.com/@csarchiquerodriguez/maquina-de-soporte-vectorial-svm-92e9f1b1b1ac>.
- Matich, Damián Jorge: *Redes Neuronales: Conceptos Básicos y Aplicaciones*. En: *Universidad Tecnológica Nacional – Facultad Regional Rosario* (2001).
- Moreira, Daniel A. y Mariana A. M. de Souza: *Teorema de Khinchin: Uma justificativa para a fórmula da entropia de Shannon*. En: *Revista Brasileira de História da Matemática* 20.42 (2020), págs. 1-17. DOI: 10.36908/rbhm.v20i42.337. <https://rbhm.org.br/index.php/RBHM/article/view/337/318>.
- Murphy, K. P.: *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

Nehme Müller, Daniel: *Máquina de Boltzmann*. En: *Universidade Federal do Rio Grande do Sul* (1993). https://danielnm.gitlab.io/inf-ufrgs/docs/t1_arqe.pdf.

Pedregosa, F. et al.: *Scikit-learn: Machine Learning in Python*. En: *Journal of Machine Learning Research* 12 (2011), págs. 2825-2830.

Q-Learning in Python. En: *GeeksforGeeks* (2021). <https://www.geeksforgeeks.org/q-learning-in-python/>.

Richardson, Mark: *Principal Component Analysis*. En: *Documento técnico* (2009). https://people.maths.ox.ac.uk/richardsonm/Principal_Component_Analysis.pdf.

Ruiz, M. German: *Modelo de análisis de datos utilizando técnicas de aprendizaje supervisado y no supervisado, para identificar patrones en la información generada por los pacientes, sometidos a juegos diseñados como un instrumento de apoyo terapéutico*. En: *Universidad Jorge Tadeo Lozano, Facultad de Ciencias Naturales e Ingeniería* (2019).

Shalizi, Cosma: *Lecture Notes on Principal Components Analysis*. En: *Carnegie Mellon University* (2012). <http://www.stat.cmu.edu/~cshalizi/350/lectures/10/lecture-10.pdf>.

Sharma, Siddharth, Simone Sharma y Anidhya Athaiya: *Activation Function in Neural Networks*. En: *International Journal of Engineering Applied Sciences and Technology* (2020).

Suthaharan, S.: *Machine Learning Models and Algorithms for Big Data Classification*. Springer, 2015.

Terrádez Gurrea, Manuel: *Análisis de Componentes Principales*. En: *Universitat Oberta de Catalunya* (2002). Proyecto e-Math, financiado por la Secretaría de Estado de Educación y Universidades (MECD). <https://www.studocu.com/ca-es/document/universitat->

oberta - de - catalunya / estadística / analysis - de - componentes - principales/69976580.

Torlai, G. y R. G. Melko: *Learning thermodynamics with Boltzmann machines*. En: *Entropy* 22.5 (2020), pág. 538. DOI: 10.3390/e22050538. <https://www.mdpi.com/1099-4300/22/5/538>.

Universidad de Sevilla: *Capítulo 4. En Proyecto final de carrera*. En: *Recurso en línea* (s.f.). <https://biblus.us.es/bibing/proyectos/use/abreproy/10807/fichero/Cap%C3%ADtulo+4.pdf>.

Waskom, Michael L.: *seaborn: statistical data visualization*. En: *Journal of Open Source Software* 6.60 (2021), pág. 3021. DOI: 10.21105/joss.03021. <https://doi.org/10.21105/joss.03021>.

Wasserman, S. y K. Faust: *Social Network Analysis: Methods and Applications*. Cambridge University Press, 2013.

Xu, R. y D. Wunsch: *Survey of clustering algorithms*. En: *IEEE Transactions on Neural Networks* 16.3 (2005), págs. 645-678.

Yeturu, K.: *Machine learning algorithms, applications, and practices in data science*. En: *Handbook of Statistics* 43 (2020), págs. 81-206.

*

ANEXOS

Anexo A. Tipos de aprendizaje supervisado y no supervisado

⁶⁷ En el aprendizaje supervisado se considera tres formas de llevarlo a cabo, que dan lugar a los siguientes aprendizajes:

- **Aprendizaje por corrección de error:**⁶⁸ Consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos a la salida de la red, es decir, en función del error cometido en la salida.

Ejemplo:

⁶⁹ El algoritmo perceptron en scikit-learn ajusta sus pesos cada vez que comete un error en la salida. Para una entrada $X = [X_1, X_2, \dots, X_r]$ con salida deseada y_i y salida obtenida \hat{y}_i , el error se calcula como:

$$\delta_i = y_i - \hat{y}_i$$

El peso B_{ij} entre la entrada X_j y la neurona de salida y_i se actualiza usando la regla de corrección de error:

$$B_{ij} \leftarrow B_{ij} + \eta \cdot X_j \cdot \delta_i$$

⁶⁷ Damián Jorge Matich: *Redes Neuronales: Conceptos Básicos y Aplicaciones*. En: *Universidad Tecnológica Nacional – Facultad Regional Rosario* (2001).

⁶⁸ Damián Jorge Matich: *Redes Neuronales: Conceptos Básicos y Aplicaciones*. En: *Universidad Tecnológica Nacional – Facultad Regional Rosario* (2001).

⁶⁹ F. Pedregosa et al.: *Scikit-learn: Machine Learning in Python*. En: *Journal of Machine Learning Research* 12 (2011), págs. 2825-2830.

y el sesgo correspondiente se ajusta como:

$$b_i \leftarrow b_i + \eta \cdot \delta_i$$

donde η es la tasa de aprendizaje. Esta regla permite que la red minimice progresivamente el error cometido, modificando los pesos sólo cuando la salida no coincide con la esperada. El modelo ajusta sus pesos sólo si comete un error de predicción, aplicando así el principio de corrección de error.

- **Aprendizaje por refuerzo:**⁷⁰ Este método es una forma de aprendizaje supervisado que opera de manera más lenta en comparación con otros enfoques. Se caracteriza por no tener acceso a un ejemplo completo del comportamiento deseado durante el entrenamiento. En otras palabras, no se especifica explícitamente la salida deseada para una entrada dada. En el aprendizaje por refuerzo, el rol del supervisor se limita a proporcionar una señal de refuerzo que indica si la salida generada por la red es adecuada o no.

Ejemplo:

⁷¹ Q-learning es un algoritmo de aprendizaje por refuerzo sin modelo que se utiliza para entrenar agentes a tomar decisiones óptimas al interactuar con un entorno. El agente aprende mediante prueba y error, observando qué acciones producen mejores recompensas. A lo largo del tiempo, construye una tabla $Q(s, a)$ que representa el valor esperado de realizar la acción a en el estado s .

El algoritmo básico sigue estos pasos:

1. Inicializar la tabla $Q(s, a)$ con valores arbitrarios.
2. Para cada episodio:

⁷⁰ Damián Jorge Matich: *Redes Neuronales: Conceptos Básicos y Aplicaciones*. En: *Universidad Tecnológica Nacional – Facultad Regional Rosario* (2001).

⁷¹ *Q-Learning in Python*. En: *GeeksforGeeks* (2021). <https://www.geeksforgeeks.org/q-learning-in-python/>.

- a) Inicializar el estado s .
- b) Mientras el episodio no termine:
 - 1) Elegir una acción a usando una política ε -greedy.
 - 2) Ejecutar la acción a , observar recompensa r y nuevo estado s' .
 - 3) Actualizar el valor de $Q(s, a)$ mediante:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right]$$

- 4) Establecer $s \leftarrow s'$.

Este método permite que el agente aprenda políticas óptimas para maximizar su recompensa acumulada en tareas secuenciales⁷².

- **Aprendizaje estocástico:**⁷³ Este método implica introducir variaciones aleatorias en los valores de los pesos de las conexiones en la red neuronal y luego evaluar su impacto con respecto al objetivo deseado y las distribuciones de probabilidad. La idea es modificar aleatoriamente los pesos y medir la energía de la red resultante. Si la energía disminuye después de la modificación, lo que indica que la red se está acercando al comportamiento deseado, se acepta el cambio. Sin embargo, si la energía no disminuye, la aceptación del cambio se basa en una distribución de probabilidades previamente establecida.

Ejemplo:

⁷⁴El Descenso por Gradiente Estocástico es un algoritmo de aprendizaje estocástico supervisado que optimiza una función de pérdida ajustando los pesos de una red neuronal

⁷² *Q-Learning in Python*. En: *GeeksforGeeks* (2021). <https://www.geeksforgeeks.org/q-learning-in-python/>.

⁷³ Damián Jorge Matich: *Redes Neuronales: Conceptos Básicos y Aplicaciones*. En: *Universidad Tecnológica Nacional – Facultad Regional Rosario* (2001).

⁷⁴ F. Pedregosa et al.: *Scikit-learn: Machine Learning in Python*. En: *Journal of Machine Learning Research* 12 (2011), págs. 2825-2830.

utilizando ejemplos individuales seleccionados aleatoriamente. Esta aleatoriedad en la selección introduce el carácter estocástico del aprendizaje.

1. Inicializar aleatoriamente los pesos B_{ij} y sesgos b_i .
2. Repetir el proceso para múltiples pasadas sobre el conjunto de entrenamiento , actualizando los pesos después de cada muestra individual:
 - a) Barajar aleatoriamente el conjunto de entrenamiento.
 - b) Para cada muestra (X, y_i) :
 - 1) Calcular la salida estimada de la red:

$$\hat{y}_i = f \left(\sum_{j=1}^r B_{ij} X_j + b_i \right)$$

2) Calcular el error: $\delta_i = y_i - \hat{y}_i$

3) Actualizar los pesos con la regla:

$$B_{ij} \leftarrow B_{ij} + \eta \cdot X_j \cdot \delta_i$$

4) Actualizar el sesgo:

$$b_i \leftarrow b_i + \eta \cdot \delta_i$$

3. Repetir hasta convergencia o alcanzar el número máximo de pasadas sobre el conjunto de entrenamiento.

El algoritmo implementado en Scikit-learn mediante `SGDClassifier` y `SGDRegressor` permite optimizar diferentes funciones de pérdida..

Ejemplo: Observe la siguiente imagen que corresponde al conjunto de datos;

En este caso, que es el aprendizaje supervisado, teniendo en cuenta el conjunto de datos y además unas etiquetas donde se enseña a la computadora que las figuras con 4 lados se llaman cuadrados y las figuras con 3 lados se llaman triángulos. Además que tenga en cuenta que ciertos

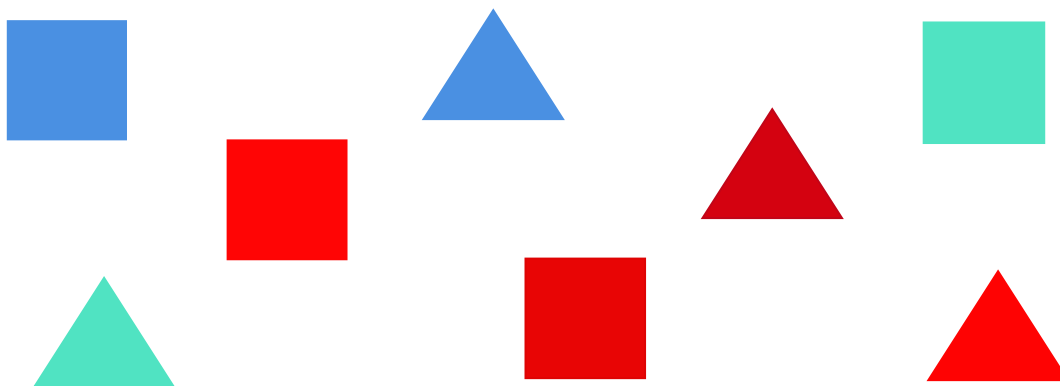


Figura 60. Conjunto de datos

parametros en los pixeles, dan un color llamado azul y otros parametros dan el color llamado rojo. Así ha aprendido a clasificar en cuadrados y triángulos azules, y en cuadrados y triángulos rojos, como podemos observar en la siguiente imagen:

En el ámbito de los algoritmos de aprendizaje no supervisado, se suelen identificar principalmente dos tipos:

- **Aprendizaje Hebbiano:**⁷⁵ Esta regla de aprendizaje, que sirve como fundamento para muchas otras, busca evaluar la familiaridad o extraer características de los datos de entrada. Su principio se basa en una suposición bastante directa: si dos neuronas, h_i y h_j , adoptan el mismo estado simultáneamente (ya sea activas o inactivas), entonces se incrementa el peso de la conexión entre ambas. Las entradas y salidas permitidas para la neurona son: $\{-1, 1\}$ o $\{0, 1\}$ (neuronas binarias). Esto se puede explicar porque la regla de aprendizaje de Hebb se originó a partir de la neurona biológica clásica, que solo puede tener dos estados: activa o inactiva.

⁷⁵ Damián Jorge Match: *Redes Neuronales: Conceptos Básicos y Aplicaciones*. En: *Universidad Tecnológica Nacional – Facultad Regional Rosario* (2001).

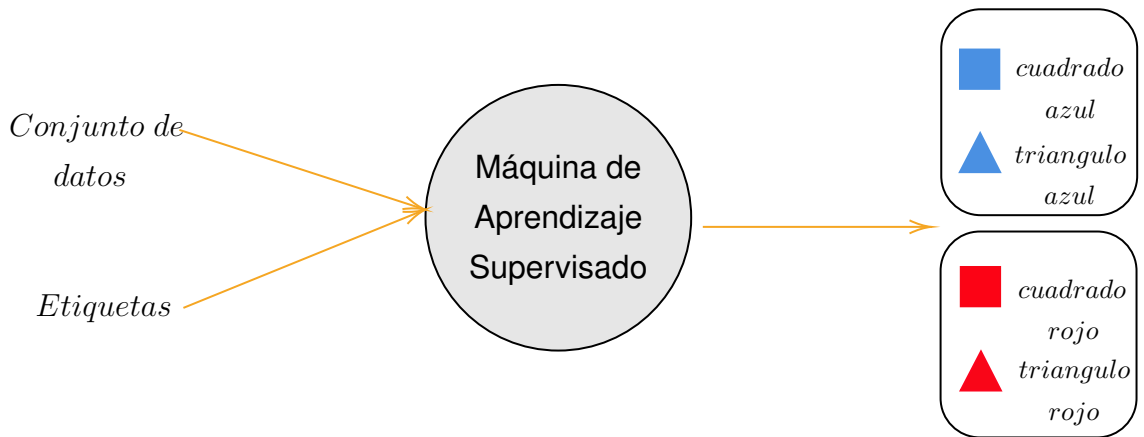


Figura 61. Aprendizaje Hebbiano

- **Aprendizaje competitivo y comparativo:**⁷⁶ Este aprendizaje se centra en la agrupación o clasificación de los datos de entrada. Una característica clave del aprendizaje competitivo es que, si se identifica un nuevo patrón como parte de una clase previamente reconocida, su inclusión ajustará la representación de esa clase. Si el patrón de entrada no se reconoce como parte de ninguna clase anteriormente identificada, entonces la estructura y los pesos de la red neuronal se ajustarán para reconocer esta nueva clase.

Ejemplo: En este caso, los algoritmos aprenden de la información dada, que es la imagen y el conjunto de datos, el aprendizaje no supervisado no va a saber que hay figuras de 4 lados ni de 3 lados, pero si sabe que en las figuras hay patrones que se repiten y así poderlas clasificar en un conjunto, lo mismo sucede con el color, de esta manera resuelve el problema:

Anexo B. Variables de interés

- **COD_DPTO:**

Departamento donde ocurrió la defunción. Hecho vital, nacido vivo o defunción, ocurrido

⁷⁶ Damián Jorge Matich: *Redes Neuronales: Conceptos Básicos y Aplicaciones*. En: *Universidad Tecnológica Nacional – Facultad Regional Rosario* (2001).

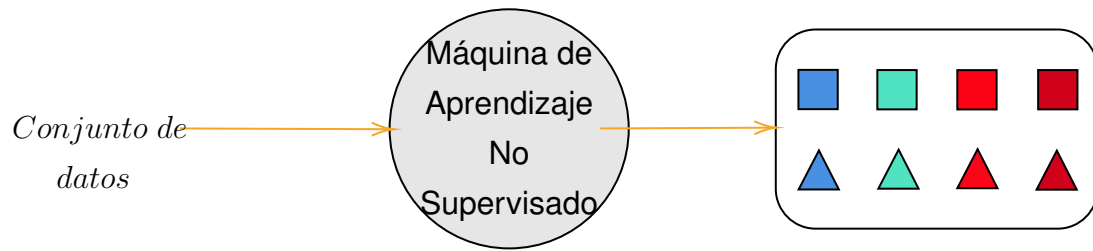


Figura 62. Funcionamiento de la red neuronal

en el territorio nacional en cada año calendario. Los siguientes números en la encuesta indican el departamento donde ocurrió la defunción: 05=Antioquia, 08=Atlántico, 11=Bogotá, 13=Bolívar, 15=Boyacá, 17=Caldas, 18=Caquetá, 19=Cauca, 20=Cesar, 23=Córdoba, 25=Cundinamarca, 27=Chocó, 41=Huila, 44=La Guajira, 47=Magdalena, 50=Meta, 52=Nariño, 54=Norte de Santander, 63=Quindío, 66=Risaralda, 68=Santander, 70=Sucre, 73=Tolima, 76=Valle del Cauca, 81=Arauca, 85=Casanare, 86=Putumayo, 88=Archipiélago de San Andrés, Providencia y Santa Catalina, 91=Amazonas, 94=Guainía, 95=Guaviare, 97=Vaupés, 99=Vichada.

Se quiere resaltar la variable departamento ya que, según el informe Instituto Nacional de Salud. Informe de Evento Mortalidad Perinatal y neonatal tardía. 2021. Colombia"(ver⁷⁷), persisten diferencias en grupos poblacionales en cuanto a la magnitud y el ritmo de la disminución del riesgo de morir durante el periodo perinatal y los primeros 28 días de vida. Es así como Chocó, Vichada, Guainía, Vaupés y La Guajira reportan las razones más elevadas siendo las entidades territoriales que representan los porcentajes más altos de personas con necesidades básicas insatisfechas (NBI) en 2018, lo que resalta su influencia en las defunciones fetales.

■ **SEG_SOCIAL:**

⁷⁷ Colombia. Mortalidad Perinatal y Neonatal Tardía. En: Instituto Nacional de Salud (2021). <https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA%20INFORME%202021.pdf>.

Régimen de seguridad social del fallecido (para muerte fetal, o de menor de un año el de la madre). Los siguientes números indican la Entidad Administradora en Salud, a la que pertenecía el fallecido (para muerte fetal): 1=Entidad promotora de salud, 2=Entidad promotora de salud - Subsidiado, 3=Entidad adaptada de salud, 4=Entidad especial de salud, 5=Entidad exceptuada de salud, 9=Sin información.

Se busca resaltar la importancia de la seguridad social, ya que se considera que tiene un impacto significativo en las tasas de defunciones fetales. Según el informe "Instituto Nacional de Salud. Informe de Evento Mortalidad Perinatal y neonatal tardía. 2020. Colombia"(ver⁷⁸), al realizar el análisis de las variables sociodemográficas se observó que las razones de mortalidad perinatal y neonatal tardía más altas se concentraron en el centro poblado (19,8 por cada 1.000 nacidos vivos), afiliados al régimen especial (74,8 por cada 1.000 nacidos vivos) y no afiliados al régimen de seguridad social en salud (19,1 por cada 1.000 nacidos vivos); y con pertenencia étnica ROM (267,9 por cada 1.000 nacidos vivos), raizal (61,2 por cada 1.000 nacidos vivos) e indígena (20,5 por cada 1.000 nacidos vivos).

■ **NIV_EDUM:**

Último nivel de estudios que aprobó la madre. Los siguientes números indican el último nivel de estudios aprobado: 1=Preescolar, 2=Básica primaria, 3=Básica secundaria, 4=Media académica o clásica, 5=Media técnica, 6=Normalista, 7=Técnica profesional, 8=Tecnológica, 9=Profesional, 10=Especialización, 11=Maestría, 12=Doctorado, 13=Ninguno, 99=Sin información.

■ **EDAD_MADRE:**

Edad de la madre del feto o del menor de un año fallecido. Los siguientes números indican la edad de la madre: 1=De 10 a 14 años, 2=De 15 a 19 años, 3=De 20 a 24 años, 4=De 25

⁷⁸ Colombia. Mortalidad Perinatal y Neonatal Tardía. En: Instituto Nacional de Salud (2020). https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA_2020.pdf.

a 29 años, 5=De 30 a 34 años, 6=De 35 a 39 años, 7=De 40 a 44 años, 8=De 45 a 49 años, 9=De 50 a 54 años, 99=Sin información.

Para estas dos últimas variables cabe destacar que, según el informe "Instituto Nacional de Salud. Informe de Evento Mortalidad Perinatal y neonatal tardía. 2021. Colombia"(ver⁷⁹), las razones más altas según la edad de la madre se encuentran en el grupo de 10 a 14 años y mayores de 40. Esta situación es similar a la identificada en otros países, donde el embarazo antes de los 20 y después de los 35 años se asocia a un mayor riesgo materno y perinatal. Se ha identificado que en las adolescentes existe un mayor riesgo biológico que puede ocasionar parto prematuro, bajo peso al nacer y restricción del crecimiento intrauterino. De igual forma, para este grupo de edad, los determinantes sociales de la salud como la escolaridad y la falta de redes de apoyo tienen una alta relevancia en el progreso de la gestación y en los resultados perinatales. Por otra parte, para las mujeres de edad avanzada el riesgo aumenta debido a la presencia de patologías propias de la gestación y una mayor frecuencia de patologías maternas crónicas.

Anexo C. Análisis complementario de la entropía por clúster

Como complemento al análisis principal de la entropía presentado en el Capítulo 3, se incluye una representación gráfica adicional mediante un mapa de calor. Esta visualización permite observar de manera individual la dispersión de cada variable socioeconómica dentro de los clústeres generados por el modelo K-means.

⁷⁹ Colombia. Mortalidad Perinatal y Neonatal Tardía. En: Instituto Nacional de Salud (2021). <https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20PERINATAL%20Y%20NEONATAL%20TARD%C3%8DA%20INFORME%202021.pdf>.

Antes

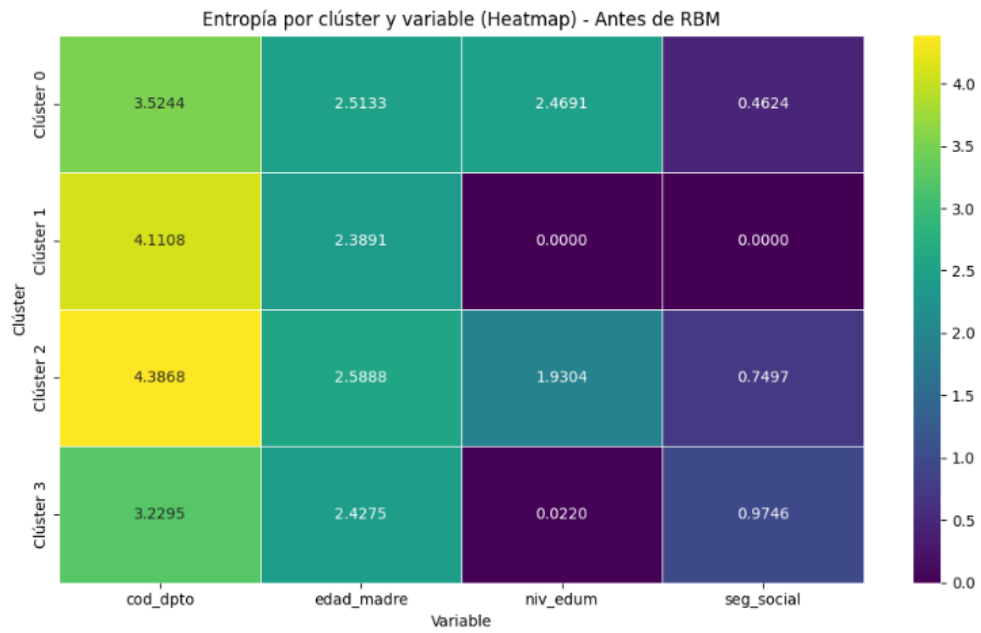


Figura 63. Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2020.

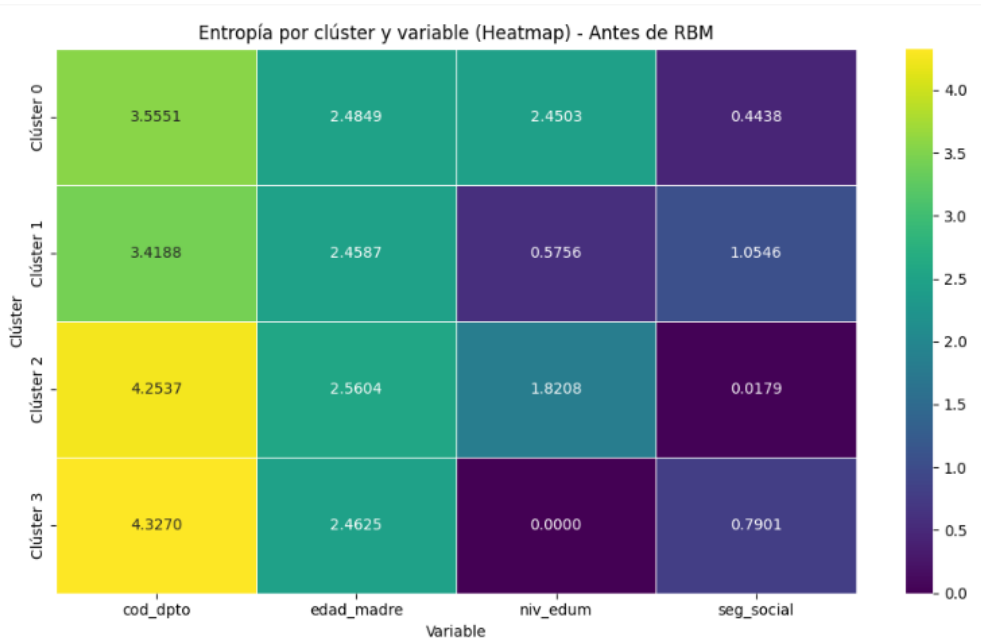


Figura 64. Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2021.

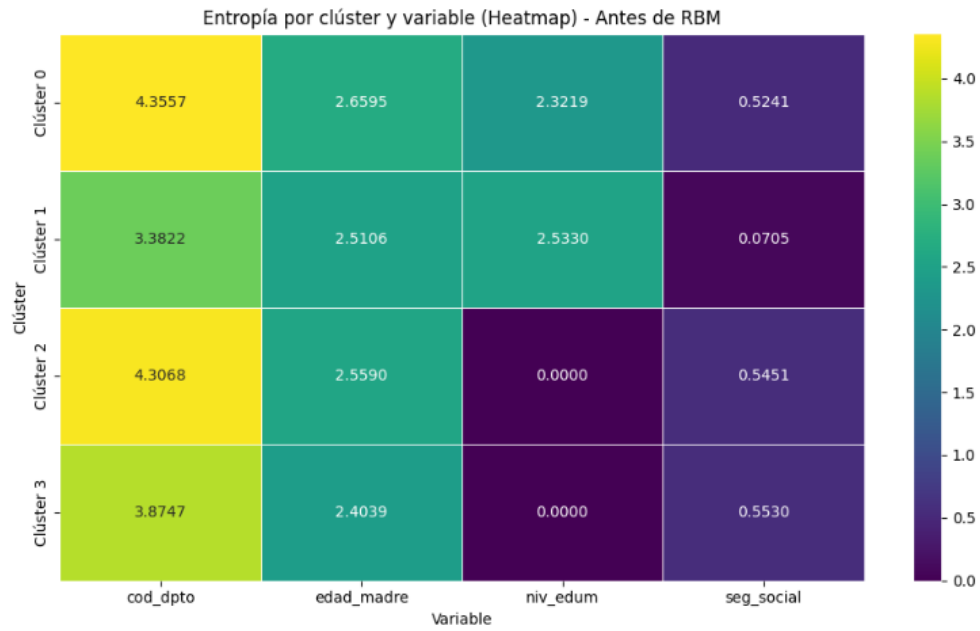


Figura 65. Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2023.

Después

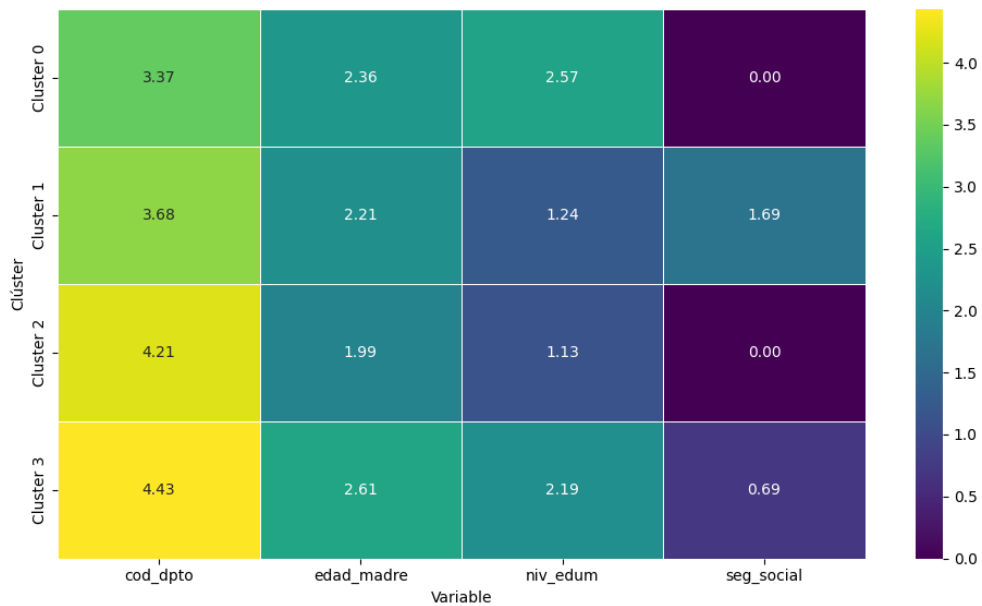


Figura 66. Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2020.

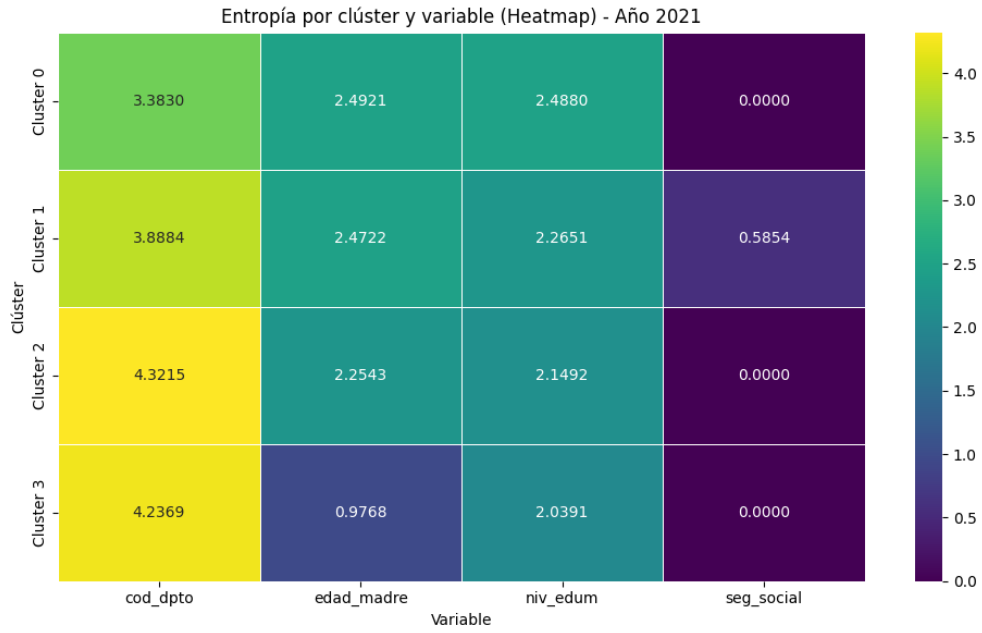


Figura 67. Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2021.

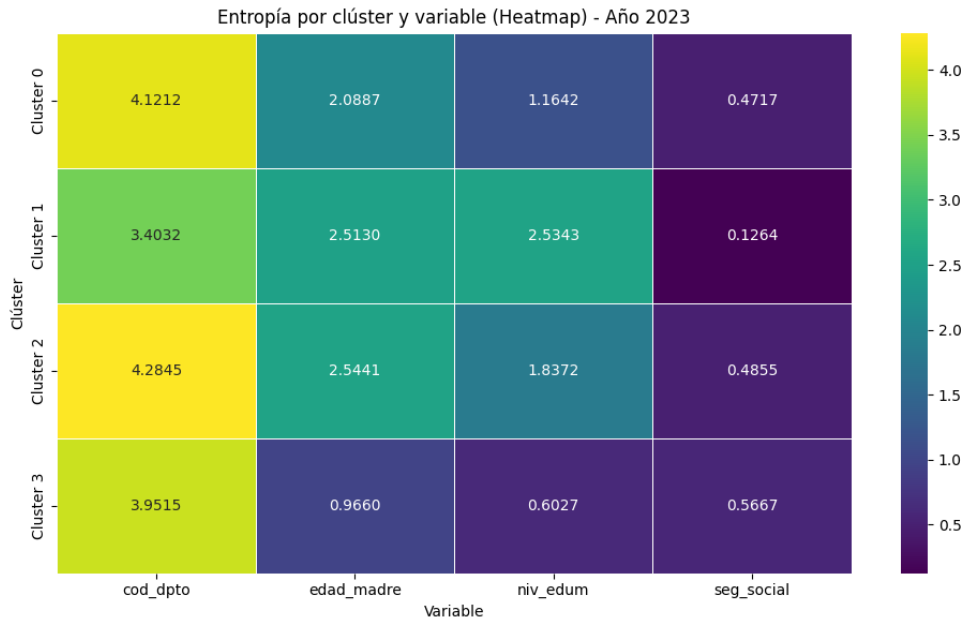


Figura 68. Mapa de calor de la entropía por clúster y variable socioeconómica. Año 2023.

Anexo D. Código

Se desarrollaron un total de seis scripts, organizados según las fases establecidas en la metodología y los años de análisis:

- **Fase “Antes”** (pre-RBM): 2020, 2021 y 2023.
- **Fase “Después”** (post-RBM): 2020, 2021 y 2023.

Todos los scripts estarán disponibles en un repositorio público de GitHub, junto con la documentación necesaria para su ejecución y reproducción de resultados.

Enlace al repositorio: [Repositorio en GitHub](#).