

Análisis de los determinantes socioeconómicos para el cumplimiento de las recomendaciones de actividad física en Colombia

David Fernando García Algarra¹, y Leidy Nathalia Gómez Díaz²

^{1,2}Fundación Universitaria Los Libertadores

Septiembre, 2023

Resumen

Contexto: La inactividad física se ha catalogado como un gran problema de salud pública a nivel global, ya que está asociado con el desarrollo de diversas enfermedades crónicas no transmisibles, como la diabetes, ciertos tipos de cáncer y la obesidad, entre otras. Por esta razón, la Organización Mundial de la Salud (OMS) recomienda participar en al menos 150 minutos de actividad física de intensidad moderada por semana durante el tiempo libre como medida para mantener una salud adecuada. **Propósito:** Identificar los principales determinantes socioeconómicos que influyen en el cumplimiento de esta directriz de la OMS, con el fin de generar recomendaciones de políticas públicas para promover la actividad física en Colombia. **Metodología:** Se llevó a cabo un análisis descriptivo y el desarrollo de un modelo clasificador de refuerzo de gradiente utilizando datos proporcionados por el Ministerio de Salud y Protección Social de Colombia, recopilados a través de la Encuesta Nacional de Situación Nutricional (ENSIN) en 2015. La muestra utilizada comprendió un total de 16,509 observaciones anonimizadas de individuos de 18 a 64 años de edad. **Resultados:** Los principales determinantes están relacionados con el género, el índice de riqueza armonizado, la disponibilidad de instalaciones deportivas en el área residencial, el uso de vehículos, la región, la educación, entre otros. **Conclusiones:** Basándose en los resultados, se sugiere que las políticas públicas prioricen la promoción de la actividad física, especialmente entre las mujeres, las personas con un estatus socioeconómico más bajo y aquellos con hábitos sedentarios. Además, es esencial considerar las diferencias regionales al diseñar e implementar estas políticas. **Palabras clave—** Actividad física, Salud Pública, Determinantes socioeconómicos, Gradient Boosting Classifier

Abstract

Context: Physical inactivity has been classified as a major global public health issue, as it is associated with the development of various non-communicable chronic diseases such as diabetes, certain types of cancer, and obesity, among others. For this reason, the World Health Organization (WHO) recommends engaging in at least 150 minutes of physical activity per week as a measure to maintain adequate health. **Purpose:** To identify the main socioeconomic determinants influencing compliance with this WHO guideline, in order to generate public policy recommendations that encourage physical activity in Colombia. **Methodology:** A descriptive analysis and the development of a gradient boosting classifier model were conducted using data provided by the Ministry of Health and Social Protection of Colombia, collected through the National Nutritional Situation Survey (ENSIN) in 2015. The sample used consisted of a total of 16,509 anonymized observations of individuals between 18 and 64 years of age. **Results:** The main determinants are related to gender, harmonized wealth index, the presence of sports facilities in the residential area, vehicle usage, region, education, among others. **Conclusions:** Public policies should prioritize the promotion of physical activity, especially among women, individuals with lower socioeconomic status, and those with sedentary habits. Additionally, it is essential to consider regional differences when designing and implementing these policies. **Keywords—** Physical Activity, Public Health, Socioeconomic Determinants, Gradient Boosting Classifier

1. Introducción

La Organización Mundial de la Salud (2022) (OMS), define la actividad física (AF) como cualquier movimiento corporal que involucra los músculos esqueléticos y consume energía. La inactividad física (IF), por otro lado, se ha convertido en uno de los principales factores de riesgo para la salud, contribuyendo significativamente a la aparición de enfermedades crónicas no transmisibles. De hecho, se ha demostrado que las personas con niveles insuficientes de AF tienen un riesgo de mortalidad entre un 20% y un 30% mayor en comparación con aquellas que mantienen niveles adecuados de actividad. Además,

se destaca que el aumento en los niveles de IF tiene efectos negativos en “los sistemas de salud, el medio ambiente, el desarrollo económico, el bienestar de la comunidad y la calidad de vida” (Organización Mundial de la Salud, 2022). Por lo tanto, la OMS, propone una serie de recomendaciones, entre las que se destaca la sugerencia de que los adultos de 18 a 64 años realicen al menos 150 minutos de actividad física por semana como una medida fundamental para mantener una buena salud.

En este orden de ideas, se precisa que la práctica de actividad física trae consigo grandes beneficios que ayudan a contrarrestar la aparición de diversas

enfermedades crónicas no transmisibles, como la obesidad, y ayuda a disminuir la prevalencia de problemas psicológicos como el estrés, depresión y ansiedad (Huang y Humphreys, 2012). Asimismo, la OMS indica que la AF regular trae beneficios positivos sobre el sistema muscular, óseo, cardiorrespiratorio, entre otros, donde también cabe destacar los beneficios para los sistemas de salud de los países (Organización Mundial de la Salud, 2022).

Los estudios relacionados con el análisis de la AF, han cobrado mayor relevancia en los últimos años, debido a la relación que se ha identificado entre la inactividad física y diversas enfermedades. En consecuencia, la mayoría de las investigaciones en este campo se han centrado en identificar los factores clave que influyen en si una persona decide llevar a cabo actividades físicas o no.

Estos determinantes suelen estar relacionados en su mayoría con variables socioeconómicas, como el ingreso monetario, la edad, el género, el nivel de educación y la ubicación geográfica, entre otras. Las metodologías de dichos trabajos se sustentan principalmente en modelos de regresión, haciendo uso de datos a escala nacional o regional; y en menor proporción, se aplican técnicas de aprendizaje automático para explorar y comprender de manera más profunda los factores que influyen en la práctica de AF.

En Latinoamérica, en países como Brasil, Bernardelli, Pereira, y Kortt (2020), examinan los factores económicos y sociales que influyen en la frecuencia de participación en AF para hombres y mujeres entre 18 y 64 años, usando un modelo de regresión logística ordenada convencional. Algo similar realizan Cansino y Gálvez (2014), para el caso de Perú, donde se estimó un modelo logístico, mediante el cual se observó que los individuos más educados tienen una mayor probabilidad de participar en actividades físicas, y que las características como el estrato y destinar tiempo a trabajar, disminuyen la probabilidad de realizar AF. Por su parte, en México, Cruz, Salazar, y Cruz Ramos (2013) realizan un estudio para la detección y dimensionamiento de los factores que inciden en la realización de AF, mediante el modelo teórico SLOTH propuesto por Cawley (2004).

En Europa y Estados Unidos, Humphreys y Ruseski (2011), desarrollan y estiman un modelo económico sobre actividad física basado en el modelo SLOTH (Cawley, 2004), el cual busca crear un marco económico que modele el comportamiento relacionado con la actividad física y la alimentación. Por otra parte, García, Lera-López, y Suárez (2011), extienden la teoría neoclásica estándar del consumidor para explicar la asignación de tiempo a la actividad física y deportes. Los autores asumen una función de utilidad de elasticidad de sustitución constante, estimando un modelo a través de una regresión aparentemente no relacionada y el procedimiento de dos pasos de Heckman. De forma similar, para el caso de Colombia, (Hernández Melo, 2015), identifica factores socioeconómicos como la educación y el género como determinantes de la AF haciendo uso de análisis estadísticos uni variados, bi variados y multivariados, pero desde una perspectiva netamente de salud pública.

Para el caso de los modelos de machine learning,

Lottata y Georganos (2023), clasifican la contribución de siete factores socioeconómicos del vecindario a la prevalencia de la inactividad física en Chicago, EE.UU., utilizando modelos de regresión de aprendizaje automático no lineal, comparando el rendimiento predictivo con redes neuronales artificiales ponderadas geográficamente, sugiriendo finalmente que la pobreza es el determinante más importante, mientras que, el espacio verde es el determinante menos importante en el aumento de la prevalencia de la AF. Por último, Cheng et al. (2021), examinan la relación entre la AF y el estado de peso, evaluando el rendimiento y el poder predictivo de un conjunto de modelos de machine learning y métodos estadísticos tradicionales, en el cual se utilizaron las estimaciones del modelo de regresión logística.

El presente estudio tiene como objetivo identificar los principales determinantes socioeconómicos del cumplimiento de la recomendación de la OMS de realizar por lo menos 150 minutos de AF, en tiempo libre, a la semana en Colombia. Para ello, se llevó a cabo un análisis estadístico descriptivo, junto con un modelo clasificador de reforzamiento de gradiente, haciendo uso de técnicas de machine learning. Esta investigación se basó en el análisis de datos de la Encuesta Nacional de Situación Nutricional (ENSIN) del año 2015, extrayendo una muestra de 16,509 observaciones anonimizadas de adultos entre las edades de 18 y 64 años. Los datos utilizados son de alcance nacional y ofrecen una representación de todo el país.

2. Antecedentes de política pública

La Organización Mundial de la Salud, establece una directriz que insta a los adultos a participar semanalmente en un mínimo de 150 minutos de AF moderada en tiempo libre, con el propósito de fomentar y preservar la salud (Organización Mundial de la Salud, 2022). En este contexto, se presenta la posibilidad de evaluar la adherencia a esta recomendación mediante el análisis de datos recopilados a través de la ENSIN (2015), la cual proporciona información actualizada acerca del estado nutricional y la salud de la población colombiana.

Ahora bien, es importante mencionar que en Colombia rige la ley 1355 de 2009, por medio de la cual se define la obesidad y las enfermedades crónicas no transmisibles asociadas a la inactividad física como una prioridad de salud pública y se adoptan medidas para su control, atención y prevención (Congreso de la República, 2019). Dado lo anterior, en varios artículos, se presentan estrategias para promover la actividad física, el transporte activo, dietas balanceadas y saludables, permitiendo ejecutar diversos programas y estrategias para mitigar el impacto que genera la inactividad física y sus derivados.

Por otra parte, a nivel internacional, países como Dinamarca, Suecia, Noruega, Finlandia, Países Bajos, Nueva Zelanda, Australia y Canadá, han implementado una serie de políticas públicas y programas para promover la actividad física y fomentar un estilo de vida activo en su población (Minsalud, 2022). Muchos países han invertido en infraestructura, educación, concientización, políticas

escolares, apoyo a la actividad física, en el lugar de trabajo y en la planificación urbana y de transporte.

A pesar de los avances logrados en Colombia en la promoción de la AF, todavía se enfrentan desafíos significativos. Estos incluyen la falta de acceso a espacios seguros para la AF en algunas regiones y la necesidad de aumentar la conciencia pública sobre la importancia de mantener un estado de salud adecuado. En este contexto, es imperativo que Colombia fomente la participación de la mayoría de su población en diversas estrategias destinadas a promover la actividad física, como se destaca en el Informe sobre la Situación Mundial de la Actividad Física 2022 (OMS, 2022).

3. Métodos

En el presente documento, se ha desarrollado una metodología compuesta por cinco pasos. En primer lugar, una revisión del estado del arte. Seguidamente, se realiza un análisis descriptivo de las variables socioeconómicas más relevantes que influyen en la adherencia de las personas a la recomendación de realizar al menos 150 minutos de actividad física. Después, se aplican las transformaciones necesarias con el fin de identificar el modelo que mejor se ajusta a los datos disponibles, en este caso se utiliza el Gradient Boosting Classifier o modelo clasificador de reforzamiento de gradiente. Posteriormente, se procede a la interpretación de los resultados con el objetivo de identificar los principales determinantes y sus efectos. Por último, se formulan recomendaciones de política pública basadas en los hallazgos obtenidos a lo largo de la investigación.

3.1. Datos

Se llevó a cabo un análisis del cumplimiento de las recomendaciones de AF, a través de la utilización de datos de la tercera ENSIN (2015). Este análisis se presenta como un elemento estadístico fundamental para el monitoreo de políticas públicas, tanto a nivel nacional como territorial, en el ámbito de la salud, alimentación y nutrición.

La tabla de datos empleada en este estudio comprende un total de 16,509 registros, abarcando 11 variables que presentan aspectos demográficos, socioeconómicos, así como hábitos y estilos de vida de los participantes. Cabe mencionar que la población de interés para esta investigación se encuentra en el rango de edades comprendido entre los 18 y 64 años en Colombia.

En este contexto, la variable dependiente categórica se formula de la siguiente manera: *si una persona cumple o no con las recomendaciones de actividad física en tiempo libre*. Esta variable será fundamental para predecir si, basándonos en diversas variables socioeconómicas, de salud y hábitos de vida, una persona logrará cumplir con el objetivo de realizar al menos 150 minutos de actividad física semanal.

En el análisis exploratorio de algunas de estas variables, se destaca que el 58.9% de los hombres y el 41.1% de las mujeres cumplen con las recomendaciones

de actividad física según los estándares de la OMS, como se ilustra en la figura 1. Además, en la figura 2, se observa que las regiones que registran un mayor cumplimiento de la meta de los 150 minutos semanales de actividad física son las regiones oriental y central, con un 19,8% y un 19%, respectivamente.

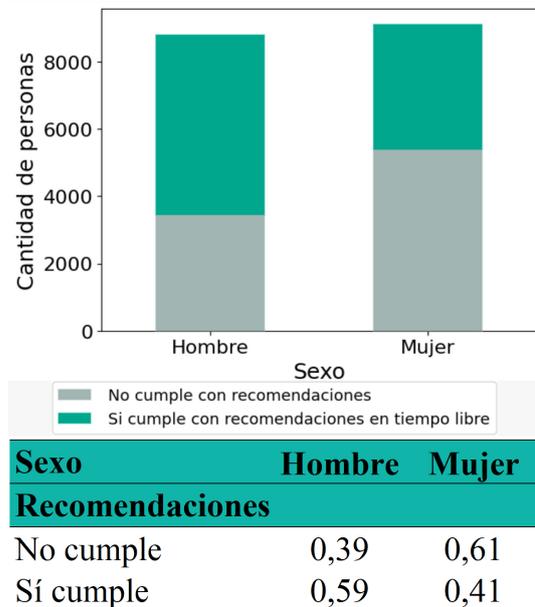


Figura 1: Cumple o no con las recomendaciones de actividad física de acuerdo al sexo de la persona.

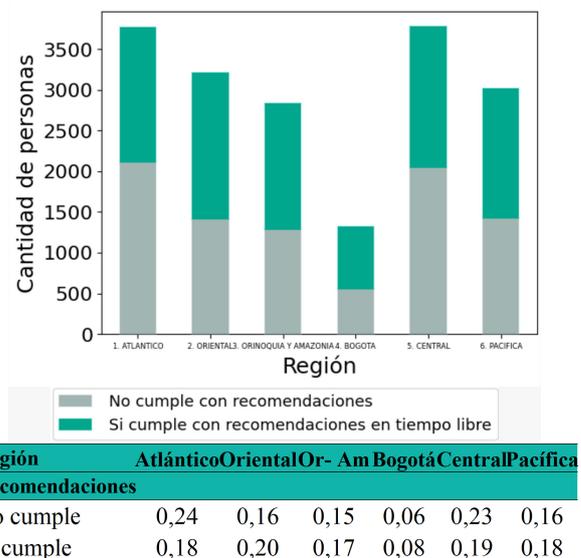
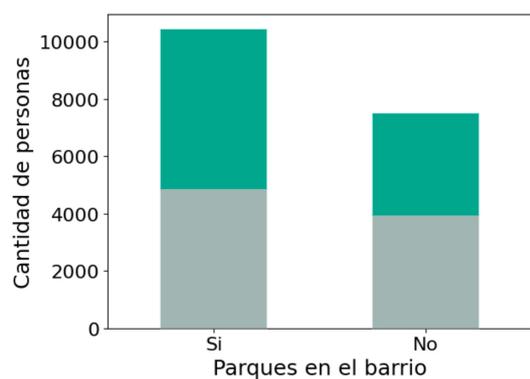


Figura 2: Cumple o no con las recomendaciones de actividad física de acuerdo a la región donde se encuentre.

Además, según lo representado en la figura 3, se evidencia que cuando el barrio o vereda cuenta con parques, áreas verdes, centros recreativos o centros deportivos, el cumplimiento de las recomendaciones alcanza un 60.9%.



Parques barrios	Sí	No
No cumple	0,55	0,45
Sí cumple	0,61	0,39

Figura 3: Cumple o no con las recomendaciones de actividad física si existen en el barrio o vereda parques, áreas verdes, centros recreativos o centros deportivo.

3.2. Exploración de datos

Después del estudio de la base de datos, que incluyó el manejo estadístico de los valores faltantes y una depuración detallada, se han elegido un conjunto de variables explicativas numéricas presentadas en la tabla 1, como la edad, el número de días de uso de computadora, teléfono inteligente e internet, el índice de riqueza armonizado, los años de educación, el índice de masa corporal y los minutos totales por semana en vehículo automotor. Además, se han considerado variables explicativas categóricas, como el género, la presencia de parques, áreas verdes, centros recreativos o deportivos en el barrio o vereda, la región, el conocimiento de los programas de actividad física dirigidos en el barrio y la afiliación a un régimen de aseguramiento en salud, como se explica en la tabla 2.

En este contexto, se han seleccionado un total de 11 variables que servirán para analizar y explicar el cumplimiento de las recomendaciones de actividad física en Colombia. Este análisis se basa en una muestra de 16,509 observaciones anonimizadas de adultos con edades comprendidas entre los 18 y 64 años.

Variable	Mean	StdDev.	Min	Max
Índice de Riqueza Armonizado	-1,01	3,41	-13,38	3,94
Edad (años)	38,82	13,44	18	64
Educación (años)	8,62	4,59	0	24
IMC	26,48	6,44	9,74	384,71
Auto semana (minutos)	355,2	674,08	0	7140
Uso internet, computador, smartphone (días)	2,85	3,28	0	7

Tabla 1: Resumen de las variables numéricas.

Variable	Descripción
Cumple las recomendaciones	0. No 1. Sí
Sexo	0. Hombre 1. Mujer
¿En su barrio hay parques, áreas verdes, centros recreativos o centros deportivos?	0. No 1. Sí
Región	1. Atlántico 2. Oriental 3. Orinoquía y Amazonía 4. Bogotá 5. Central 6. Pacífica
¿Conoce programas de actividad física dirigida en el barrio?	0. No 1. Sí
Aseguramiento en salud	1. Régimen Contributivo o Especial 2. Régimen Subsidiado 3. No afiliado

Tabla 2: Descripción de las variables categóricas.

3.3. Modificación de datos

Con el objetivo de obtener un modelo adecuado para los propósitos de esta investigación, se llevó a cabo un proceso de limpieza de datos. En este proceso se excluyeron las filas que presentaban algún valor faltante tanto para la tabla de entrenamiento como para la de prueba, con el fin de tener un conjunto de datos homogéneo. Para la variable que indica si los individuos tienen conocimiento o no de programas de actividad física dirigidos, se aplicó una técnica de imputación de datos, que consistió en asignar el valor más frecuente (moda) a las instancias con datos faltantes.

Para la codificación de las variables categóricas, se utilizó el transformador "One Hot Encoder", con la configuración de eliminar la primera categoría codificada en cada variable. Esta estrategia se adoptó con el fin de evitar la multicolinealidad y prevenir el error conocido como "dummy trap". Por otro lado, en relación a las variables numéricas, se aplicó el transformador "Standard Scaler" con el propósito de estandarizar los datos. Este proceso implica una transformación que ajusta las características numéricas para que tengan una media de cero y una desviación estándar de uno.

El preprocesamiento de datos realizado es esencial para asegurar la calidad y la idoneidad de los datos en este estudio. Involucra la limpieza de datos, la codificación de características categóricas y la normalización de características numéricas, creando así una base sólida para el análisis estadístico y el desarrollo de modelos. Estas técnicas son críticas para garantizar resultados robustos y válidos.

3.4. Modelo

Se realiza una aproximación al modelo clasificador de reforzamiento de gradiente (Gradient Boosting Classifier model), haciendo uso de la librería Pycaret. Esta técnica se emplea para desarrollar modelos de clasificación y regresión, mejorando así el proceso de aprendizaje del modelo. Los modelos de refuerzo de gradiente son inherentemente no lineales y están compuestos por árboles de decisión o regresión, que se utilizan para realizar predicciones basadas en los nodos de decisión (Lusa et al., 2017). Los árboles de regresión individuales tienden a mejorar su precisión cuando se consideran como parte de un conjunto, lo que lleva a la construcción progresiva de conjuntos para abordar errores previos (Chakrabarty, Kundu, Dandapat, Sarkar, y Kole, 2019).

Siguiendo este enfoque, el modelo de refuerzo de gradiente busca minimizar el error acumulado al combinar múltiples árboles de decisión mediante un proceso de optimización secuencial. En cada etapa, se construye un nuevo árbol que se enfoca en los errores residuales del conjunto anterior. Esta metodología gradual y adaptativa permite que el modelo mejore continuamente su capacidad predictiva a medida que se ajusta a los datos de entrenamiento, lo que lo convierte en una herramienta valiosa para predecir si una persona cumple o no con las recomendaciones de AF.

3.5. Evaluación

Utilizando los datos de prueba, con un umbral de decisión del 0.38, el modelo clasificador de refuerzo de gradiente (Gradient Boosting Classifier), obtuvo una precisión del 55.84%. Esta métrica refleja la proporción de predicciones correctas en relación con el total de predicciones realizadas por el modelo en el conjunto de prueba. La alta tasa de recuperación del 92.62% indica que el modelo es eficaz en identificar la mayoría de las instancias que pertenecen a la clase positiva (sí cumple con las recomendaciones).

La precisión del 53.98% indica la proporción de predicciones correctas específicamente en la clase positiva, lo que resulta en un puntaje F1 del 68.21%. El puntaje F1 es una medida que combina precisión y recuperación en una sola métrica y refleja un equilibrio entre ambas. El rendimiento general del modelo se considera moderado o aceptable, como lo indican el índice Kappa del 10.14% y el índice MCC (Coeficiente de Correlación de Matthews) del 15.2%. El índice Kappa mide la concordancia entre las predicciones del modelo y las observaciones reales, mientras que el índice MCC es un indicador de la calidad general de las predicciones del modelo.

En la segunda evaluación, que utiliza los datos de entrenamiento, el modelo presenta una ligera mejora en su precisión, alcanzando un valor del 56.31%. Esta mejora sugiere que el modelo se desempeña ligeramente mejor en los datos con los que fue entrenado, aunque la diferencia es marginal en comparación con la evaluación anterior. Estos datos se aprecian en la tabla 3.

Test data							
Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
GBC	0.5584	0.6558	0.926	0.54	0.68	0.1014	0.152

Train data							
Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
GBC	0.5631	0.6638	0.935	0.54	0.69	0.1108	0.169

Tabla 3: Métricas de evaluación y entrenamiento.

Por otro lado, la matriz de confusión del modelo (4), indica que hay un buen desempeño al predecir instancias de la clase 1 (cumple con las recomendaciones), acertando en 1624 casos positivos. Sin embargo, presenta algo de dificultad al predecir la clase 0 (no cumple con recomendaciones), con 1011 falsos positivos. Esto indica que el modelo podría realizar algunas predicciones erróneas en lo que respecta a la clase 0. El umbral de clasificación fue del 0.38.

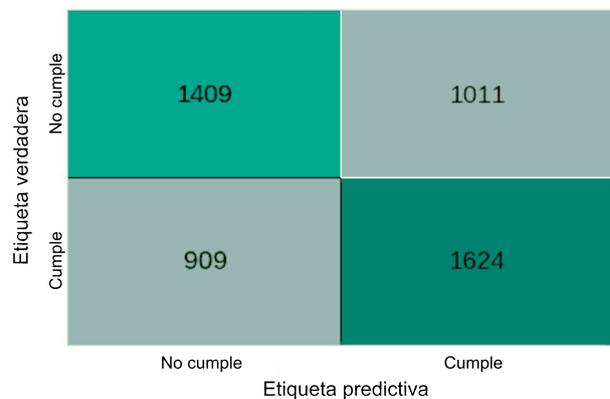


Figura 4: Matriz de confusión del modelo.

De forma similar, la curva de aprendizaje que se observa en la figura 5, indica que el modelo mejora en el entrenamiento a medida que se agregan más datos. Esto sugiere que el modelo está aprovechando la información adicional para ajustarse mejor a los patrones de los datos, lo cual es un signo positivo en términos de su capacidad de generalización. En resumen, la matriz de confusión y la curva de aprendizaje revelan la capacidad del modelo para predecir clases, destacando sus fortalezas y áreas de mejora a la hora de clasificar.

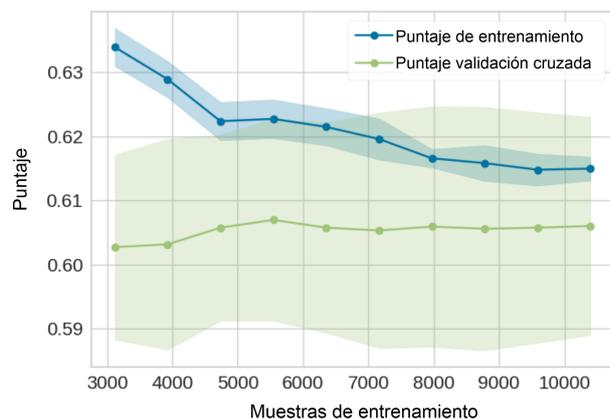


Figura 5: Curva de aprendizaje.

4. Resultados y discusión

4.1. Importancia de características

Una vez que se ha ajustado el modelo Gradient Boosting Classifier, se genera un gráfico de barras que ilustra la importancia de cada variable en relación con la predicción del cumplimiento de las recomendaciones de actividad física que indica la OMS. Dicha medida, según, Adler y Painsky (2022) intenta dar un resumen sobre la importancia de cada variable o característica para la predicción del modelo.

En este contexto, se utiliza la codificación de 0 para indicar que la persona no cumple con las recomendaciones y 1 para notar que sí las cumple. Este análisis de importancia de características es esencial en el campo del aprendizaje automático, ya que permite identificar qué variables tienen un impacto significativo en las predicciones del modelo, como se observa en el gráfico 6.

Se destaca que la variable “sexo” con la categoría “Mujer” se posiciona como la más influyente en el modelo, con una importancia asignada de 0.53. Este resultado sugiere que el género puede desempeñar un papel fundamental en la predicción del cumplimiento de las recomendaciones de actividad física. La identificación de esta variable como la más relevante podría tener implicaciones significativas para el diseño de estrategias e intervenciones dirigidas a promover la actividad física, especialmente entre las mujeres.

Además, se observa que los “minutos totales por semana en vehículo automotor” también tienen una importancia considerable, con un valor de 0.11. Esta variable está relacionada con un estilo de vida más sedentario, lo que indica que las personas que pasan más tiempo en vehículos automotores se asocian a ser menos propensas a cumplir con las recomendaciones de actividad física. Este hallazgo respalda la idea de que reducir el uso de vehículos motorizados podría ser una estrategia eficaz para fomentar la actividad física y promover un estilo de vida más saludable.

Otras variables relevantes, aunque con una importancia ligeramente menor en comparación, son el “índice de masa corporal (IMC)” con un valor de 0.08 y el “índice de riqueza armonizada” con un valor de 0.07. El IMC es un indicador de salud que refleja la relación entre el peso y la altura de una persona, y su influencia en el modelo sugiere que el estado de peso de un individuo puede tener un impacto en su predisposición a cumplir con las recomendaciones de actividad física.

En cuanto a la variable que indica si la persona tiene conocimiento de la existencia de programas de actividad física dirigida en su entorno local, como ciclovías, clases de zumba, yoga, entre otros, aunque su importancia es relativamente menor en comparación con otras variables, su inclusión en el análisis resalta la importancia de contar con opciones de actividad física accesibles en la comunidad. Este hallazgo sugiere que la conciencia y la participación en programas de actividad física locales podrían contribuir positivamente al cumplimiento de las recomendaciones de actividad física de la OMS, al ofrecer

a las personas oportunidades cercanas y variadas para mantener un estilo de vida activo.

Además, variables como la “edad”, la pertenencia a la región Oriental y Bogotá, la presencia de “parques o áreas verdes en el barrio o vereda”, así como el “número de días que se usa computador, smartphone o internet”, también tienen un impacto en la predicción del cumplimiento de las recomendaciones de actividad física. Estos resultados indican que factores demográficos, geográficos y de estilo de vida desempeñan un papel relevante en la toma de decisiones relacionadas con la actividad física.

Por último, es importante destacar que las variables “región Central”, “región Pacífica” y “afiliación a régimen subsidiado o contributivo” no muestran una importancia significativa en la predicción del modelo, como se evidencia en la tabla 4. Estos resultados pueden ser útiles para priorizar esfuerzos y recursos en estrategias de promoción de la actividad física, centrándose en las variables más influyentes identificadas en este análisis.

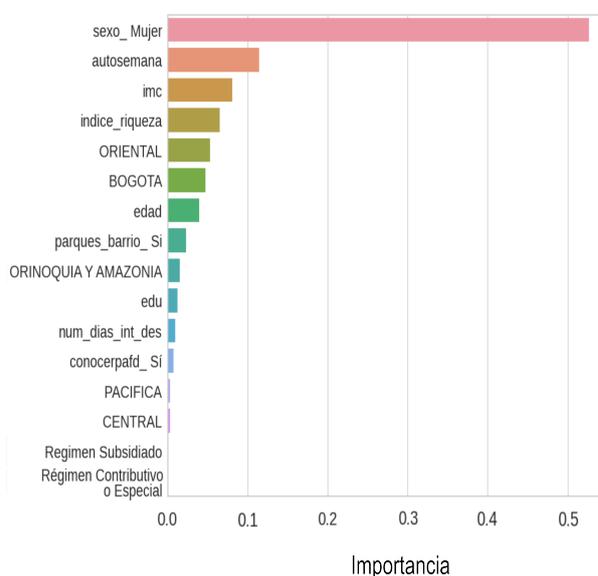


Figura 6: Gráfico de importancia de las variables.

Feature	Importance
sexo_mujer	0.53
autosemana	0.11
imc	0.08
indice_riqueza	0.07
región Oriental	0.05
región Bogotá	0.05
edad	0.04
parques_barrio_Si	0.02
región Or y Am	0.02
edu	0.01
num_dias_int_des	0.01
conocerpafd_Si	0.01
región pacífica	0.00
region central	0.00
regimen Subsidiado	0.00
régimen Contributivo	0.00

Tabla 4: Valores de las importancias de las variables en el modelo.

4.2. Análisis de valores SHAP

De acuerdo a los valores SHAP (Shapley Additive Explanations) del modelo de referencia 7, se observa que las disparidades de género representan el factor de mayor relevancia en relación con el cumplimiento de la recomendación de actividad física. En este contexto, se percibe una desventaja considerable para el género femenino, dado que el modelo tiende a clasificar predominantemente en la categoría 0 (incumplimiento de las recomendaciones) a aquellos individuos que se identifican como mujeres. Esto se debe a las dinámicas tradicionales en Colombia, donde las mujeres suelen asumir una mayor carga de responsabilidades en las tareas domésticas y el cuidado de los niños, lo que resulta en una reducción del tiempo disponible para la actividad física (Mintransporte, 2022).

Asimismo, el índice de riqueza armonizado se destaca como otro factor de relevancia, evidenciando disparidades socioeconómicas en la adhesión a esta recomendación. Concretamente, aquellos sujetos con condiciones económicas más favorables se relacionan con el cumplimiento de las directrices, en contraste con aquellos pertenecientes a estratos socioeconómicos más bajos, quienes muestran una mayor propensión a no alcanzar niveles adecuados de actividad física. Estos hallazgos se basan en la hipótesis de que las personas con ingresos y estratos económicos más bajos suelen ocupar una mayor parte de su tiempo en diversas actividades laborales, lo que limita las oportunidades de ocio y, por lo tanto, de participación en actividad física durante su tiempo libre. Similar a lo evidenciado por Cansino y Gálvez (2014), en Perú.

En cuanto a la disponibilidad de instalaciones deportivas en el entorno residencial, se observa que la presencia de infraestructuras deportivas cerca del lugar de residencia de los individuos, como parques y áreas verdes, está fuertemente asociada a una mayor probabilidad de que una persona cumpla con las recomendaciones de actividad física establecidas. Este fenómeno se refleja en el patrón discernible del modelo de evaluación, el cual muestra una clara inclinación a clasificar en la categoría 1, que denota el cumplimiento de las recomendaciones de actividad física, a aquellos individuos que disponen de tales instalaciones en su entorno cotidiano.

Este impacto se erige como un influyente impulsor que, de manera subconsciente, incita a las personas a participar activamente en actividades físicas. Este efecto, conocido en la teoría económica como “nudge”, encuentra sus fundamentos teóricos en la obra seminal de Richard Thaler y Cass Sunstein, titulada “Nudge: Improving Decisions About Health, Wealth, and Happiness” (2008). En este libro, se desarrolla la teoría del comportamiento, que postula cómo pequeñas modificaciones en el entorno pueden influir de manera positiva en las decisiones y elecciones de las personas, fomentando así comportamientos más saludables y beneficiosos, como es el caso de la actividad física regular. Este efecto

también se conoce como “Paternalismo Libertario” (Thaler y Sunstein, 2008).

De igual manera, la variable que indica si la persona conoce de programas de actividad física dirigidos en su zona de residencia, se relaciona de manera positiva con la probabilidad de cumplir con las recomendaciones, respaldando también la teoría del “nudge” y del paternalismo libertario. En este contexto, el modelo de predicción muestra una tendencia a clasificar a los individuos en la categoría de cumplimiento con las recomendaciones (clase 1) cuando estos están informados sobre la disponibilidad de programas que promueven la actividad física en su barrio o vereda. Estos programas engloban diversas iniciativas, como ciclovías, clases de zumba, yoga, entre otras.

Asimismo, para el caso de las regiones, el modelo tiende a predecir la clase 1 si la persona reside en las regiones Orinoquía, Amazonía, Oriental y Bogotá. Estas regiones parecen ofrecer un entorno más propicio para el cumplimiento de las recomendaciones, lo que puede estar relacionado con factores como la disponibilidad de espacios públicos para la práctica deportiva, la cultura de la actividad física en esas áreas o incluso aspectos socioeconómicos que influyen en las oportunidades de participación.

Por su parte, la variable que registra la cantidad total de minutos dedicados al uso de vehículo automotor a lo largo de la semana, se destaca como un indicador significativo en la predicción del cumplimiento de la pauta de actividad física, según lo revelado por el modelo. Este resultado puede explicarse por la hipótesis de que un menor uso de vehículo automotor se asocia con una mayor propensión a realizar actividades físicas. Cuando las personas dependen menos del automóvil como medio de transporte, es más probable que opten por alternativas más activas, como caminar, andar en bicicleta o utilizar el transporte público, lo que naturalmente conlleva un mayor nivel de actividad física.

Finalmente, es importante resaltar el papel desempeñado por la variable que refleja la cantidad de años de educación en el contexto del modelo predictivo. A pesar de que su efecto se manifiesta de manera modesta, es notorio que tienda a predecir la clase 0 en situaciones donde los años de educación son limitados. Este hallazgo subraya la relevancia intrínseca de la educación en la formación de la conciencia y la percepción en relación con la salud. La educación no solo proporciona conocimiento sobre la importancia de mantener un estilo de vida saludable, sino que también puede influir en la toma de decisiones informadas sobre la actividad física y otros comportamientos relacionados con la salud.

Investigaciones previas han demostrado de manera constante cómo la educación puede actuar como un determinante social de la salud, ya que las personas con niveles más altos de educación tienden a adoptar comportamientos más saludables y a tener un mayor conocimiento sobre la prevención de enfermedades, lo que en última instancia puede contribuir a una mayor adherencia a las recomendaciones de actividad física. Como lo indicó (Hernández Melo, 2015) al utilizar datos de la ENSIN 2010, donde personas con un mayor nivel

educativo suelen tener mejor nivel económico y por ende, más tiempo para dedicar al ocio y a la actividad física.

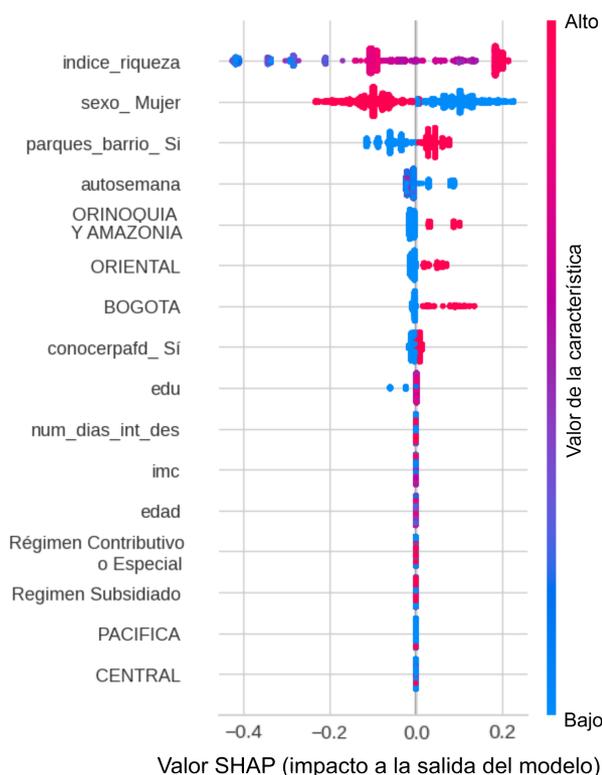


Figura 7: Shap values.

4.3. Recomendaciones de política pública

De acuerdo al análisis de importancia de las características, los valores SHAP del modelo y los resultados descriptivos, se delinean a continuación una serie de recomendaciones de política pública y salud pública destinadas a mejorar el cumplimiento de la recomendación de actividad física de la OMS, especialmente enfocándonos en los grupos más afectados identificados en los resultados presentados anteriormente:

1. Índice de Riqueza Armonizado

- Infraestructuras deportivas accesibles: Asegurar que las áreas recreativas, parques y espacios públicos para la práctica de deportes estén disponibles y sean accesibles en las comunidades con bajos niveles de riqueza. Esto podría incluir la construcción y el mantenimiento de instalaciones deportivas y áreas verdes.
- Programas de actividad física adicionales: Desarrollar programas de actividad física adicionales o gratuitos que estén disponibles para todas las edades en áreas con bajos ingresos, como clases de ejercicios grupales, actividades deportivas comunitarias y eventos recreativos.
- Educación y conciencia: Implementar programas educativos en las escuelas y comunidades para aumentar la conciencia sobre la importancia de la actividad física para la salud y brindar orientación sobre cómo incorporarla en la vida diaria.

- Apoyo a la participación: Ofrecer subsidios o becas para que los niños y jóvenes de familias con bajos ingresos puedan participar en actividades deportivas organizadas, como equipos deportivos o clubes.

- Seguridad: Mejorar la seguridad en las áreas donde se realizan actividades físicas al aire libre para que las personas se sientan cómodas al hacer ejercicio, especialmente en comunidades con altos índices de criminalidad.

2. Sexo

- Programas específicos para mujeres: Incluir clases de ejercicio adaptadas a las necesidades de las mujeres y campañas de concientización sobre la importancia de la actividad física en su salud y bienestar.
- Entornos amigables para las mujeres: Mantener una constante labor de concientización en los entornos domésticos, mediante iniciativas que comprendan la creación y rehabilitación de espacios públicos, como parques y zonas deportivas, que permitan a las familias realizar actividades físicas de forma gratuita y accesible.

3. Presencia de parques o áreas verdes

- Inversión en infraestructura: Invertir en la creación y renovación de parques y áreas verdes en todas las comunidades para garantizar que las personas de todos los estratos socioeconómicos tengan acceso a lugares adecuados para la actividad física.

4. Minutos en vehículo automotor

- Promoción del ciclismo: Implementar políticas orientadas a promover el ciclismo como una modalidad de transporte sostenible, incluyendo la construcción de carriles exclusivos para bicicletas, estacionamientos seguros para bicicletas y campañas informativas sobre los beneficios del ciclismo como alternativa de movilidad.

5. Regiones:

- Eventos deportivos y recreativos: Incentivar y ejecutar con mayor frecuencia eventos deportivos y recreativos a nivel local, regional y nacional para fomentar la participación en actividades físicas y deportivas, como carreras, caminatas, cicloturismo y actividades al aire libre.
- Adaptación a la diversidad cultural: Evaluar las condiciones particulares de cada región, teniendo en cuenta las costumbres, creencias y prácticas propias que permitan adaptar los programas de actividad física a las necesidades y recursos específicos de cada territorio, incluyendo las comunidades indígenas, afrocolombianas y raizales.

Las recomendaciones de política pública y salud pública buscan abordar las barreras identificadas en el cumplimiento de las recomendaciones de actividad física. Al enfocarse en la equidad de género, el acceso equitativo a instalaciones deportivas y la promoción del ciclismo, se puede crear un entorno propicio para que todas las personas, independientemente de su género o nivel socioeconómico, puedan disfrutar de los beneficios de un estilo de vida activos y saludables. Estas no solo contribuirán a mejorar la salud de la población, sino que también pueden tener un impacto positivo en la calidad de vida y en la reducción de las desigualdades medidas en la salud.

5. Conclusiones

A partir de los resultados descriptivos y del modelo predictivo desarrollados en este estudio, se han identificado los principales determinantes socioeconómicos del cumplimiento de las recomendaciones de la OMS, de realizar al menos 150 minutos de actividad física a la semana en tiempo libre. La aplicación de técnicas de aprendizaje automático no solo permitió explorar la importancia relativa de diversas variables, sino también comprender cómo estas variables se asocian a la probabilidad de cumplir con las directrices de actividad física.

En primer lugar, el índice de riqueza armonizado revela brechas socioeconómicas, ya que las personas con menores recursos propenden a ser más vulnerables a la hora de cumplir con las recomendaciones de la OMS. Además, se destaca el sexo como uno de los principales determinantes en el cumplimiento de las recomendaciones de actividad física. Las mujeres enfrentan una brecha en comparación con los hombres, lo que sugiere la necesidad de esfuerzos de política pública para cerrar esta desventaja de género.

Asimismo, se identifica que las políticas deben estar orientadas a combatir los estilos de vida sedentarios, como la limitación del tiempo de uso de vehículos, ya que esta variable se relaciona a la predicción con mayor probabilidad del cumplimiento de las recomendaciones cuando los minutos de uso de vehículo son bajos.

La presencia de infraestructura deportiva en el entorno residencial, el conocimiento de programas de actividad física dirigidos y la región de residencia, se han revelado como factores críticos en la determinación del cumplimiento de las recomendaciones de la OMS en cuanto a la actividad física. La cercanía de parques y

áreas verdes no solo proporciona un espacio accesible para el ejercicio, sino que también actúa como un estímulo intrínseco que motiva la participación en actividades físicas.

Es importante destacar que los programas de actividad física dirigida y la disponibilidad de escenarios deportivos actúan como un impulso suave hacia hábitos más saludables, a través del paternalismo libertario. Estos hallazgos también deben ser considerados a la luz de las diferencias regionales en Colombia, ya que cada región ofrece distintos entornos y opciones para realizar actividad física.

La implementación de las recomendaciones mencionadas anteriormente, respaldadas por la evidencia científica de diversas disciplinas, como la economía, estadística, medicina y otras áreas relacionadas, tiene el potencial de reducir la prevalencia de enfermedades crónicas no transmisibles y patologías psicológicas. Esto, a su vez, contribuirá a mejorar la calidad de vida de las personas y salvaguardar los recursos del sistema de salud nacional.

Para que las políticas sean efectivas, es fundamental involucrar a la comunidad y a las partes interesadas en su diseño e implementación. La participación activa de las personas afectadas por estas políticas puede garantizar que se adapten a las necesidades y realidades locales. Se debe establecer un sistema de seguimiento y evaluación para medir el impacto de las políticas implementadas. Esto implica la recopilación periódica de datos sobre la actividad física, la salud de la población y otros indicadores relevantes. La retroalimentación constante permitirá ajustar las políticas según sea necesario y garantizar que estén logrando sus objetivos.

Finalmente, se recomienda realizar un análisis de causalidad, con el fin de identificar de manera más precisa el impacto de las variables estudiadas en el presente documento, en relación con el cumplimiento de las recomendaciones de actividad física de la OMS.

Agradecimientos

Los datos utilizados provienen de la Encuesta Nacional de Situación Nutricional (ENSIN) 2015 y son propiedad del Ministerio de Salud y Protección Social de Colombia, del Instituto Colombiano de Bienestar Familiar y las demás instituciones implicadas en el ejercicio estadístico. Agradecemos al Ministerio por poner a disposición del público las bases de datos de la encuesta.

Referencias

- Adler, A. I., y Painsky, A. (2022). Feature importance in gradient boosting trees with cross-validation feature selection. *Entropy*, 24(5), 687.
- Bernardelli, L. V., Pereira, C., y Kortt, M. A. (2020). The economic and social determinants of participation in physical activity in Brazil. *International Review of Applied Economics*, 34(2), 252-266. Descargado de <https://doi.org/10.1080/02692171.2019.1685953>
- Cansino, K., y Gálvez, H. (2014, 01). Determinantes de la participación en actividades físicas en el Perú. *Revista Peruana de Medicina Experimental y Salud Pública*, 31, 151 - 155. Descargado de <http://www.scielo.org>

- .pe/scielo.php?script=sci_arttext&pid=S1726-46342014000100022&nrm=iso
- Cawley, J. (2004). An economic framework for understanding physical activity and eating behaviors. *American Journal of Preventive Medicine*, 27(3, Supplement), 117-125. Descargado de <https://www.sciencedirect.com/science/article/pii/S0749379704001515> doi: <https://doi.org/10.1016/j.amepre.2004.06.012>
- Chakrabarty, N., Kundu, T., Dandapat, S., Sarkar, A., y Kole, D. K. (2019). Flight arrival delay prediction using gradient boosting classifier. En *Emerging technologies in data mining and information security: Proceedings of iemis 2018, volume 2* (pp. 651–659).
- Cheng, X., Lin, S.-y., Liu, J., Liu, S., Zhang, J., Nie, P., ... Xue, H. (2021). Does physical activity predict obesity—a machine learning and statistical method-based analysis. *International Journal of environmental research and public Health*, 18(8), 3966.
- Congreso de la República. (2019). *Ley 1355 de 2009*. Bogotá: Disposiciones analizadas por Avance Jurídico Casa Editorial Ltda.
- Cruz, R., Salazar, J., y Cruz Ramos, A. (2013). Determinantes de la actividad física en México. *Estudios Sociales. Revista de Alimentación Contemporánea y Desarrollo Regional*, 21, 186-209. Descargado de <https://www.redalyc.org/articulo.oa?id=41725649008>
- García, J., Lera-López, F., y Suárez, M. J. (2011). Estimation of a structural model of the determinants of the time spent on physical activity and sport: Evidence for Spain. *Journal of Sports Economics*, 12(5), 515-537. Descargado de <https://doi.org/10.1177/1527002510387080> doi: 10.1177/1527002510387080
- Hernández Melo, M. A. (2015). *Posición socioeconómica y su relación con actividad física en tiempo libre : análisis secundario a partir de la encuesta nacional de la situación nutricional en Colombia, ensin 2010* (Tesis/Trabajo de grado - Monografía - Maestría, Pontificia Universidad Javeriana, Bogotá, Colombia). Descargado de <http://hdl.handle.net/10554/16710>
- Huang, H., y Humphreys, B. R. (2012). Sports participation and happiness: Evidence from US microdata. *Journal of Economic Psychology*, 33(4), 776-793. Descargado de <https://www.sciencedirect.com/science/article/pii/S0167487012000402> doi: <https://doi.org/10.1016/j.joep.2012.02.007>
- Humphreys, B. R., y Ruseski, J. E. (2011). An economic analysis of participation and time spent in physical activity. *The B.E. Journal of Economic Analysis Policy*, 11(1). Descargado 2023-07-30, de <https://doi.org/10.2202/1935-1682.2522> doi: doi:10.2202/1935-1682.2522
- Lotfata, A., y Georganos, S. (2023). Spatial machine learning for predicting physical inactivity prevalence from socioecological determinants in Chicago, Illinois, USA. *Journal of Geographical Systems*, 1–21.
- Lusa, L., y cols. (2017). Gradient boosting for high-dimensional prediction of rare events. *Computational Statistics & Data Analysis*, 113, 19–37.
- Ministerio de Salud y Protección Social y el Instituto Colombiano de Bienestar Familiar (ICBF) y otras entidades. (2015). *Encuesta nacional de situación nutricional (ensin) - 2015*. Survey. Descargado de <https://enlinea.minsalud.gov.co/Encuestas/Microdatos.aspx?E=ENSIN2015> (Recuperado en [2023] de [<https://enlinea.minsalud.gov.co/Encuestas/Microdatos.aspx?E=ENSIN2015>])
- Minsalud. (2022). *Orientaciones para la gestión territorial de la promoción de la actividad física*. Bogotá: Autor.
- Mintransporte. (2022). *Estrategia nacional de movilidad activa con enfoque de género y diferencial*. Bruselas, Bélgica: Unión Europea.
- OMS. (2022). *Informe sobre la situación mundial de la actividad física 2022: resumen ejecutivo*. Switzerland: Departamento de Promoción de la Salud Organización Mundial de la Salud.
- Organización Mundial de la Salud. (2022). *Actividad física*. <https://www.who.int/es/news-room/fact-sheets/detail/physical-activity>.
- Thaler, R. H., y Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.