



**TÍTULO DEL ESTUDIO: “MACHINE LEARNING PARA LA
PREDICCIÓN DE SOBREPESO Y OBESIDAD EN PACIENTES
ATENDIDOS POR EL PROGRAMA DE CONTROL DE
CRECIMIENTO Y DESARROLLO DE LA RED ASISTENCIAL
SABOGAL DEL SEGURO SOCIAL DE SALUD DEL PERÚ”**

REPORTE DE RESULTADOS DE INVESTIGACIÓN 05-2023

JUAN ALBERTO SANTILLANA CALLIRGOS
Director de IETSI

DAYSÍ DÍAZ OBREGÓN
Gerente de la Dirección de Investigación en Salud – IETSI

Autores

- Paulo Vela Antón
- Percy Soto Becerra
- Karen Eliana Bonilla Aguilar
- Martina Guillermo Roman
- Moisés Apolaya Segura

Reporte de resultados de investigación 05-2023

El presente reporte es el resultado de una investigación realizada en el marco de los temas de investigación en salud prioritarios para ESSALUD, para el periodo 2020-2022, aprobados con la Resolución de IETSI N° 150-IETSI-ESSALUD-2023.

Tema de salud en investigación prioritario: Malnutrición.

Conflicto de intereses

Los responsables de la elaboración del presente documento declaran no tener ningún conflicto de interés financiero o no financiero, con relación a los temas descritos en el presente documento.

Aprobación Ética

Este estudio fue aprobado para su ejecución por el Comité Institucional de Ética en Investigación del Instituto Nacional de Salud con fecha 30 de marzo de 2022.

Financiamiento

Este documento técnico ha sido parcialmente financiado por el Instituto de Evaluación de Tecnologías en Salud e Investigación (IETSI), ESSALUD, Perú y por el Centro Nacional de Salud Ocupacional y Protección del Ambiente del Instituto Nacional de Salud del Perú, MINSA.

Citación

Este documento debe ser citado como: “Machine Learning para la predicción de sobrepeso y obesidad en pacientes atendidos por el Programa de Control de Crecimiento y Desarrollo de la Red Asistencial Sabogal del Seguro Social de Salud del Perú”. Reporte de resultados de investigación 05-2023. Lima: ESSALUD; 2023”.

Datos de contacto

Paulo Vela Anton

Correo electrónico: paulovelamech@gmail.com

Teléfono: (+511)265 6000, anexo 1966

Contenido

Resumen	6
Introducción	7
Métodos	9
Resultados	13
Discusión y Conclusiones	17
Referencias Bibliográficas	19
Material Suplementario	21

RESUMEN

OBJETIVOS: Desarrollar y validar un modelo de Machine Learning para predecir el desarrollo de sobrepeso y obesidad infantil en pacientes atendidos por el programa de Control de Crecimiento y Desarrollo (CRED) de centros de atención primaria de la Red Asistencial de Sabogal del Seguro Social de Salud del Perú durante el mes de enero del 2019 a diciembre del 2021.

MÉTODOS: Estudio de cohorte retrospectiva de seguimiento desde los 6 meses hasta los 2 años. La población de estudio estará conformada por pacientes sanos menores de 2 años que fueron atendidos por el programa CRED durante el mes de enero del 2019 a diciembre del 2021 en la Red Asistencial Sabogal. Por otro lado, se incluirán variables sociodemográficas, clínicas, comorbilidades y datos de la madre. Se solicitó los datos pertinentes a la Gerencia Central de Tecnologías de Información y Comunicaciones (GCTIC) correspondientes al programa CRED de los Centros Asistenciales Periféricos (CAP) y Policlínicos de la Red Asistencial de Sabogal. Dichos datos fueron preprocesados para la fase de entrenamiento de modelos de Machine Learning. Luego, se realizó el entrenamiento de diversos modelos de Machine Learning y un análisis de las diferentes métricas de rendimiento con el objetivo de determinar el modelo de predicción más confiable.

RESULTADOS: EL modelo dinámico basado en épocas mediante TensorFlow y Keras obtuvo una AUC de 0,90 (0,001), superando a todos los demás modelos. También logró un mejor rendimiento estadísticamente significativo que todos los demás modelos en métricas para un clasificador estándar: precisión 30,90 % (0,22 %), F1-score 44,60 % (0,26 %), precisión 66,14 % (0,41 %) y especificidad 63,27 % (0,41 %).

CONCLUSIONES: Los modelos de predicción de la obesidad en paciente del Programa de Crecimiento y Desarrollo de EsSalud se desarrollaron a partir de datos retrospectivos, tomando un marco de tiempo para la predicción de la incidencia de la obesidad hasta los 3 años de edad. El flujo de trabajo de desarrollo del modelo de aprendizaje automático presentado se puede adaptar a varios estudios basados en datos de la historia clínica electrónica y puede ser valioso para desarrollar otros modelos de predicción clínica con registros de pacientes latinoamericanos.

Palabras clave: obesidad infantil, CRED, machine learning, inteligencia artificial

INTRODUCCIÓN

La transición nutricional, es decir, el cambio de las dietas tradicionales por alimentos más ricos en grasas, carnes y azúcar, y el aumento de los estilos de vida sedentarios a medida que los países se van industrializando, se ha asociado con el desarrollo económico, según se establece en un modelo conceptual publicado recientemente (1). Esta transición nutricional está dividida en etapas según la variación del sistema alimentario que se presenta a lo largo de los años. América Latina se encuentra en la segunda etapa de transición, mostrando una brecha cada vez más reducida de la obesidad entre las diferencias socioeconómicas (1,2). La obesidad y sobrepeso infantil se ha incrementado dramáticamente a partir de 1990, según la base de datos de la OMS (3). Existen diversos estudios no congruentes sobre los posibles factores causales, como la inseguridad alimentaria (4–6) y/o el incremento de la comida chatarra (7,8). Existe una fuerte evidencia que apoya la efectividad de las intervenciones escolares para la prevención de sobrepeso y obesidad infantil (9). Pero, no toda intervención ha demostrado tener un impacto significativo, esto va depender del tipo de intervención y los componentes de apoyo disponibles, por ejemplo las intervenciones de tipo mixto que incluyen la dieta y la actividad física han demostrado mejores efectos (10,11). En Perú, se realizan intervenciones preventivas utilizando la promoción de lactancia materna, vigilancia del crecimiento del niño y la promoción de alimentación complementaria adecuada (12).

Si bien se conocen los factores modificables del sobrepeso y la obesidad, es difícil predecir en la primera infancia. En los últimos años se han estudiado los modelos matemáticos para la predicción de obesidad (13). Con la base de datos de Wirral y la cohorte de niños del milenio se ha logrado realizar una predicción de sobrepeso y obesidad a los 3 años de edad considerando un registro de hasta dos años y un año respectivamente (14,15); mientras que, con la base de datos de la cohorte ALSPAC (Avon Longitudinal Study of Parents and Children) y CHICA (Child Health Improvement Computer Automation) se logró predicciones de hasta 5 y 10 años, con un registro de datos de hasta los 2 años de edad (16,17). Incluso, con un registro de hasta los primeros 8 años de edad, se desarrollaron modelos predictivos de sobrepeso y obesidad en adolescentes de 13 a 16 años (18). Comparando todos estos modelos predictivos, el que obtuvo una mejor precisión fue este último, con una sensibilidad y especificidad de 76% y 83% respectivamente (19). Sin embargo, aún no se ha podido determinar el mejor modelo predictivo ya que depende mucho del conjunto de datos que se tiene, pero se ha podido apreciar que los modelos de aprendizaje automático han demostrado tener un mejor rendimiento a comparación de otros modelos, uno de los más prometedores son los algoritmos de redes neuronales (13).

En el Perú, el Seguro Social de Salud (EsSalud) incorporó el Sistema de Gestión de Salud denominado EsSI (Servicio de Salud Inteligente) durante el 2019, con la finalidad de digitalizar las historias clínicas (20). Parte de las historias clínicas son los controles de crecimiento y desarrollo (CRED), los cuales se realizan de forma constante a lo largo de los primeros 4 años de vida, en puntos de corte establecidos. Luego, estos controles se realizan esporádicamente hasta los 12 o 17 años (21). Actualmente, se puede observar un incremento en la prevalencia de sobrepeso y obesidad en menores de 5 años a partir del 2015 (22). Prevenir que siga aumentando y mitigar la prevalencia tendrá un impacto futuro en las enfermedades crónicas, la esperanza de vida, salud mental, factores sociales y rendimiento tanto físico como mental (23,24). En el mundo, ya se están estableciendo estrategias de intervención temprana para la prevención de obesidad infantil y demostrando su eficacia en los primeros años de vida (25). Por lo planteado anteriormente, nuestro objetivo es realizar la identificación temprana de riesgo de obesidad infantil, desarrollando un modelo predictivo a partir de los datos recopilados en los controles de crecimiento y desarrollo de los policlínicos y centros de atención primaria del Seguro Social de Salud.

MÉTODOS

Diseño del estudio

Estudio de cohorte retrospectiva de seguimiento desde los 6 meses hasta los 2 años. La población de estudio estará conformada por pacientes sanos menores de 2 años que fueron atendidos por el programa CRED durante el mes de enero del 2019 a diciembre del 2021 en la Red Asistencial Sabogal. Por otro lado, se incluirán variables sociodemográficas, clínicas, comorbilidades y datos de la madre.

Población

La población de estudio estará conformada por infantes menores de dos años de edad atendidos con regularidad en los establecimientos de salud de la Red Sabogal durante el mes de enero del 2019 a diciembre del 2021.

Procedimientos

El presente estudio no realizó un muestreo, en cambio, se tomó a todos los participantes que cumplieron con los criterios de selección de la cohorte de desarrollo. Sin embargo, para asegurarnos de que el número de instancias sea suficientemente grande para una adecuada validación interna, se realizó una estimación del tamaño de muestra mínimo necesario. Fórmulas para el cálculo del tamaño de muestra en estudios de desarrollo y validación de algoritmos aún no han sido bien establecidas por consenso (19-21), por lo que se utilizó las recomendaciones dadas por diversos autores basándose en estudios de simulación, las cuales indican que se deberían tener al menos 10 eventos (Obesidad) y 10 no eventos (peso normal) y óptimamente entre 10 y 30 eventos y no eventos por variable incluida en el algoritmo (22-24). Se contó con al menos 100 eventos y 100 no eventos.

De acuerdo con la información otorgada por el Sistema Estadístico de Salud (SES), indican que, al año, entre el 9% - 10% de los infantes menores de tres años atendidos por consultas de pediatría desarrollan exceso de peso. Así, asumiendo un escenario conservador de observar una proporción de obesidad infantil de 9% en todos los establecimientos de salud que serán objetos de estudio, se recopiló datos retrospectivos de al menos 1112 pacientes para tener un número apropiado y óptimo de eventos ($n = 100$) y no eventos ($n = 1012$). Teniendo en cuenta la pérdida de información por datos incompletos y asumiendo que esta podría llegar a ser del 20%, se enroló retrospectivamente a 1335 infantes que cumplan los criterios de selección descritos. Como se observa, el número de infantes registrados disponible es mayor al n mínimo estimado, mejorando esto la calidad de predicción del algoritmo desarrollado.

Variables respuesta

Los desenlaces principales fueron la presencia de sobrepeso y obesidad. Los diagnósticos fueron obtenidos a partir de la base de datos de la historia clínica electrónica. El sobrepeso infantil fue definido en base a las tablas de crecimiento de la OMS, indicando un percentil de IMC por edad igual o mayor al percentil 85 hasta por debajo del percentil 95. Por último, obesidad infantil se define para pacientes que indican un percentil de IMC por edad igual o mayor al percentil 95.

Covariables

Seleccionamos tres puntos de corte clave para la detección de sobrepeso u obesidad respecto a la edad del paciente: 6 meses, 8 meses y a los 24 meses. Los puntos de corte se aplicaron para variables como peso, talla, IMC, ancho de pliegue tricípital, etc.

Manejo de bases de datos y Análisis estadístico

Las bases de datos crudas (raw data) anonimizadas fueron proporcionadas por GCTIC y las áreas involucradas en la investigación. Estas fueron procesadas para generar una base de datos limpia para el desarrollo y validación interna. Ambas fueron manejadas por el investigador principal y los coinvestigadores durante la fase preprocesamiento y modelamiento.

Durante la fase de preprocesamiento, se realizó un control de calidad de los datos ingresados para verificar la existencia de inconsistencias, valores no plausibles u otros problemas. Al encontrarse tales problemas, se recurrió a la fuente original (historia clínica electrónica) para su verificación o alguna fuente de datos auxiliar que permita su validación de acuerdo con reglas lógicas preestablecidas. En los casos que no fue posible subsanar la inconsistencia encontrada, se procedió a considerar como perdido el dato. Este proceso de control de calidad de datos fue iterativo hasta eliminar inconsistencias, dando por concluido el proceso de limpieza para almacenar y preparar las bases de datos para el modelamiento. Las bases de datos crudas y finales (ambas anonimizadas) fueron guardadas en el espacio de almacenamiento en la nube del investigador principal, teniendo respaldos de esta en la computadora de un coinvestigador de IETSI. Todo el proceso de limpieza de datos se realizó en los programas SQL-server, RStudio y Jupyter Notebook.

Desarrollo del Modelo de Aprendizaje Automático

1. Ingeniería y transformación de variables

A partir de las 15 variables administrativas y clínicas, se generaron indicadores de diferentes secciones. Además, se realizó una imputación para los datos perdidos, es decir, si cabe una interpretación clínica razonable de alguna variable, se transformó en una variable categoría “dummy” (variable que solo toma dos posibles valores: 0 o 1) o politómica, utilizando un punto de corte establecido en la literatura. Caso contrario, se modeló la variable numérica como tal.

2. Selección de variables

Se realizó una preselección de 15 variables predictoras segmentada según los puntos de corte mencionados y en base al conocimiento clínico de este campo, con el objetivo de tener modelos que sean interpretables y con mayor probabilidad de ser generalizables. Se agregaron otras variables sintéticas no capturadas por la HCE de EsSalud para incrementar la precisión del modelo, incluyendo en total 99 variables relacionadas con la obesidad infantil. Luego a este conjunto de variables preseleccionadas se les aplicó un método de selección de características apropiadas para cada algoritmo que utilizaremos.

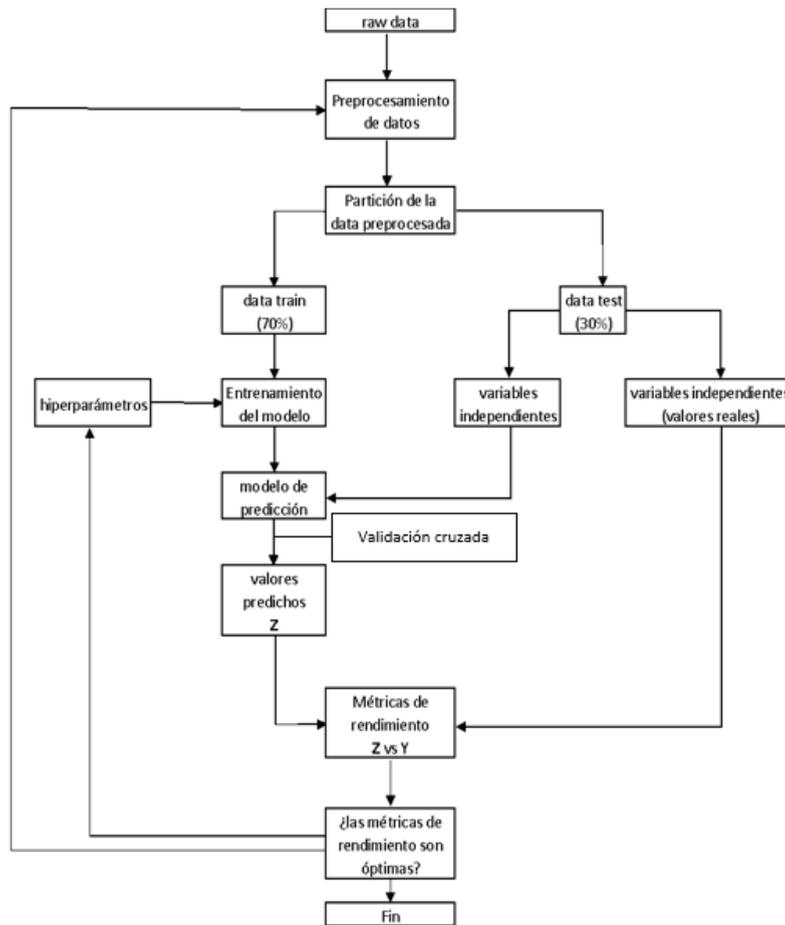
3. División de la base de datos y control de calidad

Valores biológicamente no plausibles fueron identificados y excluidos del análisis. Se obtuvieron aproximadamente 640,000 registros fueron incluidos a partir de la extracción de una base de datos de GCTIC EsSalud entre los años 2016 y 2021, dado que cumplieron con los criterios de inclusión. Tras el preprocesamiento y múltiples filtros para la calidad de datos, se obtuvo aproximadamente 16,000 pacientes menores de 2 años pertenecientes a la Red Prestacional Sabogal que cumplían con los criterios de inclusión. La base de datos se dividió siguiendo un esquema de validación cruzada anidada, en 80% para el entrenamiento, y 20% para validación. La descripción demográfica de los datos se muestra en la Tabla X. En los casos donde las métricas no fueron óptimas en la fase de validación del modelo, se procedió a realizar un balanceo de datos. Se entrenó los algoritmos sin balancear los datos y también balanceándose mediante: oversampling y se eligió el algoritmo que tenga el mejor rendimiento.

4. Entrenamiento de modelos y evaluaciones estadísticas

Se realizó el entrenamiento de 5 modelos de Machine Learning con las variables seleccionadas usando el data train para predecir obesidad. Los modelos probados fueron Decision Tree (DT), Logistic

regression (LR), Neural Network (NN), Support Vector Machine (SVM) y Extreme Gradient Boosting (XGB). Se utilizaron computadoras personales para dicho entrenamiento. Todos los modelos fueron entrenados y evaluados usando Python's Scikit Learn, haciendo una validación cruzada usando hasta 10 pliegues. Se aplicó métricas de rendimiento correspondientes a cada modelo de Machine Learning utilizado. Por ejemplo, para el caso de una regresión logística, se determinó la matriz de confusión, la exactitud (Accuracy), la sensibilidad (Recall), la precisión (Precision), la puntuación f1 (f1-Score), el área bajo la curva ROC (AUROC) y el coeficiente de Gini. Para evaluar la concordancia del conjunto de datos respecto al azar se obtuvo los valores Kappa de Cohen, y para mediar la diferencia entre los valores reales y los valores predichos se halló el Coeficiente de Correlación de Matthews. El proceso realizado se detalla en el siguiente diagrama:



Aspectos éticos

El proyecto fue presentado al comité de ética en investigación del Hospital Alberto Sabogal Sologuren para su revisión y aprobación. No se solicitó el consentimiento informado de los pacientes pues no se contó con datos personales que revelen la identidad de los sujetos de estudio. Se realizó una revisión de registros clínicos virtuales los cuales no vulneran la integridad social, psicológica ni física de los participantes incluidos. Si bien se excluyeron pacientes con comorbilidades, esto se realizó debido a que son poblaciones de menor tamaño cuyo desarrollo se manifiesta de manera diferente, por lo que se requieren algoritmos específicos y estudios particulares para estos grupos de pacientes. Finalmente, se solicitó el permiso correspondiente a GCTIC, manteniendo la confidencialidad de la información de las bases de datos. Los autores declaramos no tener conflicto de interés.

RESULTADOS

Característica de la población de estudio

En total, hubo 20,000 participantes incluidos en el estudio. De estos, el 63% fue mujer y aproximadamente la mitad de ellos tenían menos de 10 meses de edad, como se observa en el gráfico de distribución de la población por edad. Asimismo, se observa una distribución homogénea entre la distribución de densidad de los registros por edad.

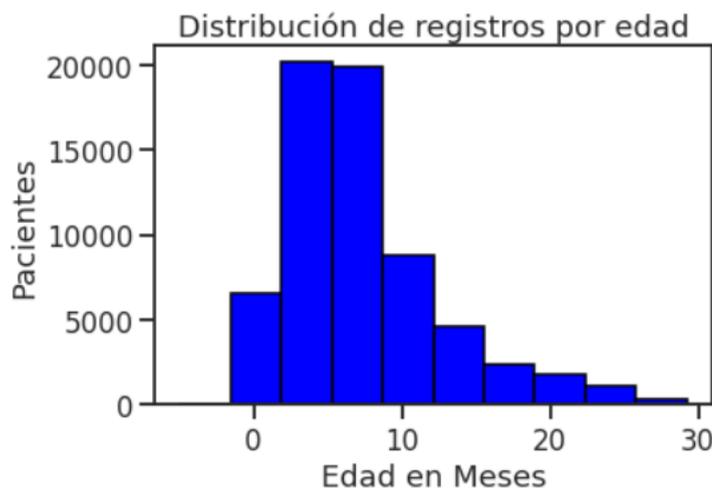


Fig. 1. Distribución de pacientes según edad en meses.

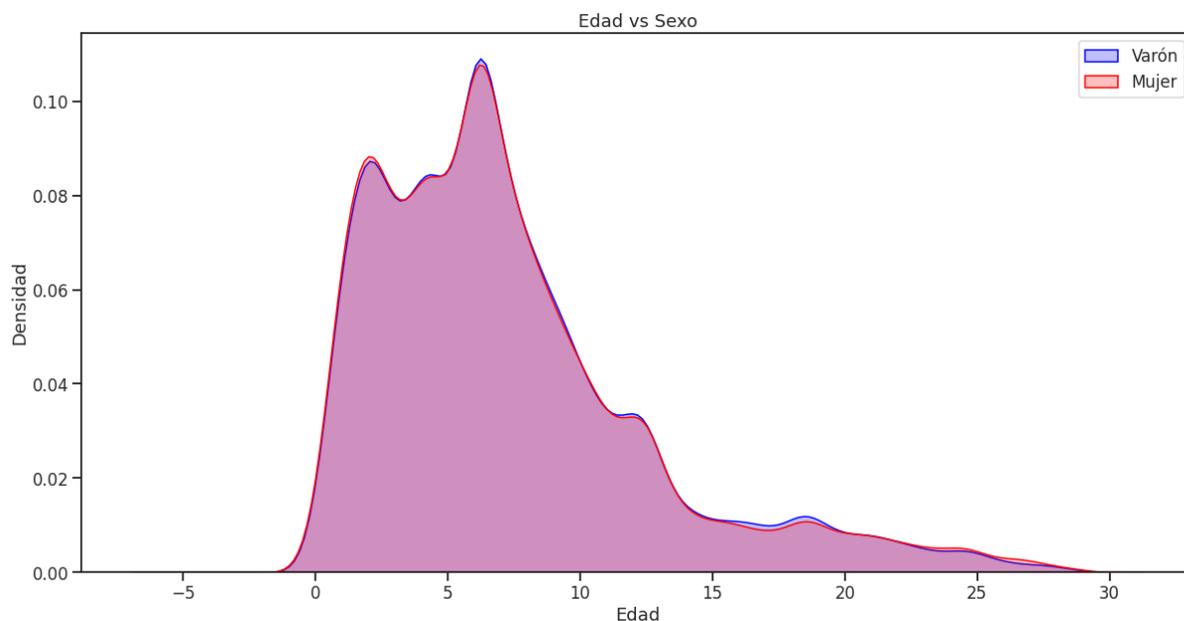


Fig. 2. Densidad de registros sexo vs edad.

Tras visualizar la distribución, se aplicaron los primeros modelos de predicción, obteniendo los siguientes resultados. La Tabla 1. muestra las características del dataset.

Etiqueta	Característica	Training	Test	Total			
0	Sobrepeso	6105	1526	7631			
1	Obesidad	2214	554	2768			
2	Sano	5071	1268	6339			
	Total de Registros del Dataset			16738			
	Tiempo de Adquisición			3 meses			
	Total de Registros del Dataset			640,900			
	Total de tamaño de la data			97.1 Mb			

	Categorías de Clasificación			3			
Support Vector Machine Algorithm							
Category	n(Truth)	N (Classified)	Precisión	F1	Recall	Accuracy	Overall Accuracy
Sobrepeso	3037	6105	0.51	0.48	0.49	49.18%	47.23%
Obesidad	1058	2214	0.49	0.50	0.49	48.39%	
Sanos	9295	5071	0.44	0.48	0.46	48.02%	
K-Nearest Neighbors Algorithm							
Category	n(Truth)	N (Classified)	Precisión	F1	Recall	Accuracy	Overall Accuracy
Sobrepeso	3048	6105	0.41	0.48	0.46	42.55%	45.81%
Obesidad	1037	2214	0.47	0.45	0.49	44.72%	
Sanos	9305	5071	0.44	0.48	0.46	45.11%	
Artificial Neural Network Algorithm							
Category	n(Truth)	N (Classified)	Precisión	F1	Recall	Accuracy	Overall Accuracy
Sobrepeso	3027	6105	0.43	0.45	0.47	41.35%	42.48%
Obesidad	1062	2214	0.40	0.42	0.47	43.82%	
Sanos	9301	5071	0.39	0.41	0.46	44.19%	

Tras las primeras pruebas con los registros, la performance de los modelos iniciales brindaba una predicción ineficiente debido a la falta de variables relevantes para el pronóstico de obesidad, por lo que se procedió a aplicar un modelo dinámico basado en Tensor Flow y Keras para evaluar un nueva performance, esta vez considerando otras variables sintéticas agregadas para conocer cuánto

incrementará la precisión. Este nuevo modelo fue entrenado a pérdida, y realiza un análisis por épocas para conocer si se dará la aparición de obesidad, seleccionado la época con mejor predicción.

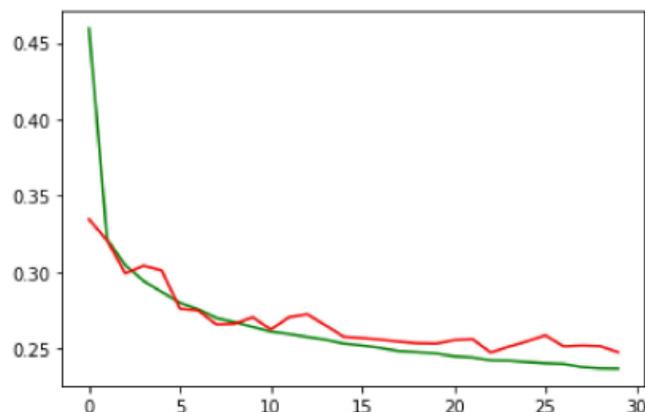


Fig. 3. Performance a pérdida de épocas

En el gráfico, observamos como el valor absoluto de la pérdida (en rojo), disminuye con el avance de las épocas. En verde se visualiza la recta de ajuste de dichos valores.

Finalmente, al seleccionar la época con mejor performance, se obtuvo un valor de predicción de 0.9102, es decir aproximadamente 91% de precisión en la predicción de la obesidad infantil en menores de 2 años. El código de ejecución puede verse a detalle en el Anexo 1.

DISCUSIÓN Y CONCLUSIONES

En este trabajo de investigación, se analizó los datos de casi 16,000 para desarrollar una cohorte de análisis para el estudio de la predicción de la obesidad infantil. Se describió en detalle un flujo de trabajo integral de análisis de datos que incluye control de calidad de datos, transformación de datos, imputación de datos faltantes y desarrollo de modelos de aprendizaje automático. Aplicamos este flujo de trabajo para identificar y excluir valores biológicamente inverosímiles, convertir registros longitudinales en variables de ventana adecuadas para la entrada del modelo y desarrollar siete modelos de aprendizaje automático para predecir el riesgo de obesidad desde el nacimiento hasta la edad <2 años. La base extraída no contaba con suficientes variables para generar una predicción de alta performance, por lo que algunas variables sintéticas tuvieron que agregarse, como pliegue tricípital, para obtener una mejor predicción. Las aplicaciones de Big Data y Machine Learning en el cuidado de la salud exhiben un gran potencial para respaldar la gestión de la salud de la población y la toma de decisiones clínicas. Sin embargo, el uso secundario de los datos del ESSI sigue siendo difícil debido a problemas de calidad de los datos, como datos incompletos, incoherencias e inexactitudes.

Además, la disponibilidad de etiquetas y la heterogeneidad de los tipos de datos dificultan la creación de modelos analíticos precisos a través del aprendizaje automático. Nos enfrentamos a todas estas preocupaciones y el flujo de trabajo presentado aquí intenta abordarlas sistemáticamente. Para aumentar la disponibilidad de etiquetas, aplicamos un estándar de la OMS calculado a partir de la altura y el peso registrados en el ESSI para determinar la obesidad. Se filtraron los valores biológicamente inverosímiles y se probaron varios métodos de imputación de datos faltantes para garantizar que los valores imputados no alterarán sustancialmente las distribuciones de datos. El rendimiento del mejor modelo, XGB, se compara favorablemente con los estudios previos limitados de predicción individualizada de la obesidad en la primera infancia mediante el aprendizaje automático, al tiempo que agrega extensiones importantes. Específicamente, nuestro modelo logró una precisión aproximada de 0,91 equivalente al valor informado para niñas y ligeramente superior al valor para niños de otros estudios. Al mismo tiempo, nuestro modelo está generalizado a niños y niñas por lo que solo se requiere un único modelo. Además, nuestros modelos predicen la obesidad hasta los 2 años. Curiosamente, este modelo se derivó de una cohorte de 16,000 individuos, pero solo logró un rendimiento similar a otros modelos reportados. Esto puede sugerir que no se pueden obtener mejoras adicionales en la predicción de la obesidad en la primera infancia simplemente aumentando el tamaño de la cohorte o la capacidad del modelo.

La interpretación del modelo es importante para comprender los factores de riesgo asociados con la obesidad infantil. Dado que la raza y el origen étnico están estrechamente correlacionados y ambos análisis calificaron la raza entre las 5 características principales, sospechamos que la diferencia de calificación en el origen étnico se debe en gran medida a cómo se manejaron las características correlacionadas. Otros factores pronósticos, como la información de los padres , los registros de nacimiento, los registros de gestación, los comportamientos sedentarios [63], los hábitos dietéticos futuros , los cambios en los factores ambientales también se sabe que son importantes para predecir la obesidad infantil. Sin embargo, debido a la limitación de nuestra fuente de datos, dicha información no se incluye en este estudio.

REFERENCIAS BIBLIOGRÁFICAS

1. Jaacks LM, Vandevijvere S, Pan A, McGowan CJ, Wallace C, Imamura F, et al. The Obesity Transition: Stages of the global epidemic. *Lancet Diabetes Endocrinol*. 2019 Mar 1;7(3):231–40.
2. Popkin BM, Reardon T. Obesity and the food system transformation in Latin America. *Obes Rev*. 2018;19(8):1028–64.
3. de Onis M, Blössner M, Borghi E. Global prevalence and trends of overweight and obesity among preschool children. *The American Journal of Clinical Nutrition*. 2010 Nov 1;92(5):1257–64.
4. Hopping BN, Erber E, Mead E, Sheehy T, Roache C, Sharma S. Socioeconomic indicators and frequency of traditional food, junk food, and fruit and vegetable consumption amongst Inuit adults in the Canadian Arctic. *J Hum Nutr Diet*. 2010 Oct;23 Suppl 1:51–8.
5. Wiles NJ, Northstone K, Emmett P, Lewis G. ‘Junk food’ diet and childhood behavioural problems: results from the ALSPAC cohort. *Eur J Clin Nutr*. 2009 Apr;63(4):491–8.
6. Jackson DB, Johnson KR. Does breast-feeding reduce offspring junk food consumption during childhood? Examinations by socio-economic status and race/ethnicity. *Public Health Nutr*. 2017 Jun;20(8):1441–51.
7. den Bosch SO, Duch H. The role of cognitive stimulation at home in low-income preschoolers’ nutrition, physical activity and body mass index. *BMC Pediatr*. 2017 Aug 1;17(1):178.
8. Popkin BM, Corvalan C, Grummer-Strawn LM. Dynamics of the double burden of malnutrition and the changing nutrition reality. *The Lancet*. 2020 Jan 4;395(10217):65–74.
9. Wang Y, Cai L, Wu Y, Wilson RF, Weston C, Fawole O, et al. What childhood obesity prevention programmes work? A systematic review and meta-analysis. *Obes Rev*. 2015 Jul;16(7):547–65.
10. Waters E, de Silva-Sanigorski A, Hall BJ, Brown T, Campbell KJ, Gao Y, et al. Interventions for preventing obesity in children. *Cochrane Database Syst Rev*. 2011 Dec 7;(12):CD001871.
11. Bahia L, Schaan CW, Sparrenberger K, Abreu G de A, Barufaldi LA, Coutinho W, et al. Overview of meta-analysis on prevention and treatment of childhood obesity. *J Pediatr (Rio J)*. 2019 Aug;95(4):385–400.
12. Bonilla C, Hajar G, Márquez D, Aramburú A, Aparco JP, Gutiérrez EL. Intervenciones para prevenir la aparición de sobrepeso y obesidad en niños menores de cinco años. *Revista Peruana de Medicina Experimental y Salud Pública*. 2017 Dec 2;34(4):682–9.
13. Rautiainen I, Äyrämö S. Predicting overweight and obesity in later life from childhood data: A review of predictive modeling approaches. *arXiv:191108361 [cs, stat] [Internet]*. 2019 Nov 19 [cited 2022 Feb 18]; Available from: <http://arxiv.org/abs/1911.08361>
14. Zhang S, Tjortjis C, Zeng X, Qiao H, Buchan I, Keane J. Comparing data mining methods with logistic regression in childhood obesity prediction. *Inf Syst Front*. 2009 Sep 1;11(4):449–60.

15. Weng SF, Redsell SA, Nathan D, Swift JA, Yang M, Glazebrook C. Estimating overweight risk in childhood from predictors during infancy. *Pediatrics*. 2013 Aug;132(2):e414-421.
16. Dugan TM, Mukhopadhyay S, Carroll A, Downs S. Machine Learning Techniques for Prediction of Early Childhood Obesity. *Appl Clin Inform*. 2015;6(3):506–20.
17. Redsell SA, Weng S, Swift JA, Nathan D, Glazebrook C. Validation, Optimal Threshold Determination, and Clinical Utility of the Infant Risk of Overweight Checklist for Early Prevention of Child Overweight. *Child Obes*. 2016 Jun;12(3):202–9.
18. Gravversen L, Sørensen TIA, Gerds TA, Petersen L, Sovio U, Kaakinen M, et al. Prediction of adolescent and adult adiposity outcomes from early life anthropometrics. *Obesity (Silver Spring)*. 2015 Jan;23(1):162–9.
19. Rautiainen I, Kauppi J-P, Ruohonen T, Karhu E, Lukkarinen K, Äyrämö S. Predicting Future Overweight and Obesity from Childhood Growth Data: A Case Study. In: Tuovinen T, Periaux J, Neittaanmäki P, editors. *Computational Sciences and Artificial Intelligence in Industry: New Digital Technologies for Solving Future Societal and Economical Challenges* [Internet]. Cham: Springer International Publishing; 2022 [cited 2022 Feb 18]. p. 189–201. (Intelligent Systems, Control and Automation: Science and Engineering). Available from: https://doi.org/10.1007/978-3-030-70787-3_13
20. EsSalud W. EsSalud implementa historia clínica digital para atención de asegurados | EsSalud [Internet]. [cited 2022 Feb 18]. Available from: <http://www.essalud.gob.pe/essalud-implementa-historia-clinica-digital-para-atencion-de-asegurados/>
21. EsSalud. Documento Técnico Orientador: Atención del Crecimiento y Desarrollo de la niña, niño y adolescente en el contexto de la COVID-19, en el Seguro Social de Salud-EsSalud [Internet]. 2021. Available from: http://www.essalud.gob.pe/downloads/doc_tecnico_orientador_CRED_17_8_VF.pdf
22. Dávila WV, Vargas RSV. Informe Gerencial SIEN HIS. 2021;46.
23. Daniels SR. The consequences of childhood overweight and obesity. *Future Child*. 2006;16(1):47–67.
24. Se W, Ji G. Childhood Overweight and Obesity: Affecting Factors, Education and Intervention. *J Child Obes* [Internet]. 2018 [cited 2022 Feb 18];03(02). Available from: <http://childhood-obesity.imedpub.com/childhood-overweight-and-obesity-affecting-factors-education-and-intervention.php?aid=22485>
25. World Health Organization. Health in 2015: from MDGs, Millennium Development Goals to SDGs, Sustainable Development Goals [Internet]. Geneva: World Health Organization; 2015 [cited 2022 Feb 25]. Available from: <https://apps.who.int/iris/handle/10665/200009>

MATERIAL SUPLEMENTARIO

ANEXO 1. [CÓDIGO DE MODELO PREDICTIVO](#)