

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**



**Máster en Big Data y Data Science: ciencia e ingeniería de datos**

## **TRABAJO FIN DE MÁSTER**

**PREDICCIÓN DE MALNUTRICIÓN INFANTIL  
UTILIZANDO PYSPARK Y APRENDIZAJE AUTOMÁTICO**

**Merve Celik Cakiroglu**  
**Tutor: Paulo Angel Villegas Nuñez**

**Septiembre 2024**



# **PREDICCIÓN DE MALNUTRICIÓN INFANTIL UTILIZANDO PYSPARK Y APRENDIZAJE AUTOMÁTICO**

**AUTOR: Merve Celik Cakiroglu alumna**  
**TUTOR: Paulo Angel Villegas Nuñez tutor**

**Escuela Politécnica Superior**  
**Universidad Autónoma de Madrid**  
**Septiembre de 2024**





# Resumen

Los problemas de salud afectan a todos los países del mundo, y estos problemas están siendo monitoreados y se están tomando medidas por la Organización Mundial de la Salud (OMS), reconocida por muchos países. Por ejemplo, el Covid, que ha afectado recientemente a todo el mundo, es uno de los ejemplos más comunes. La malnutrición infantil es otro tema de preocupación para la OMS.

La malnutrición en niños se puede definir como el crecimiento y desarrollo anormal de los niños debido a comer muy poca o demasiada comida y tener irregularidades nutricionales. En este tema, la OMS clasifica a los niños con cuatro problemas diferentes: retraso del crecimiento, debilitante, bajo peso y sobrepeso, utilizando parámetros corporales de los niños.

A pesar de una disminución en el número de niños afectados por la desnutrición, sigue siendo una de las principales causas de muerte, representando el 45% de las fatalidades infantiles. Por lo tanto, comprender y pronosticar la malnutrición infantil es esencial para mejorar los resultados de salud infantil. El Machine Learning, que brinda soluciones a problemas complejos que no pueden ser resueltos por humanos solos, es una herramienta que se puede utilizar en este sentido.

Por lo tanto, diferentes modelos de clasificación de Machine Learning (ML) fueron entrenados y probados con PySpark. Los datos se recopilaron del Programa DHS, donde diferentes países de diferentes continentes proporcionan resultados de encuestas, ya que la OMS no suministra datos brutos para usar en modelos de ML.

Después de determinar las características importantes con Random Forest y probar los modelos uno por uno, se logró una precisión máxima del 70.8%. En el entrenamiento de los modelos, se utilizaron primero etiquetas multiclase y binarias con menor éxito. Luego, se eliminaron del conjunto de datos tres características con datos insuficientes, identificadas con la matriz de confusión, y se logró el mayor éxito solo con las muestras de retraso del crecimiento y normales.

Se espera que el estudio y sus resultados guíen futuros estudios y proporcionen soluciones más precisas a los problemas que se puedan encontrar en la investigación sobre la malnutrición infantil.



## ***Agradecimientos***

En primer lugar, quiero expresar mi más sincero agradecimiento a mi tutor, Paulo Villegas. Él me guió, compartió su experiencia conmigo y siempre respondió rápidamente a todas mis preguntas. Su conocimiento y apoyo han sido clave para mi interés en Spark y para la calidad de mi tesis.

También quiero dar las gracias a todos mis profesores. Su dedicación, pasión y enseñanzas han despertado en mí un gran interés por el aprendizaje automático y me han dado los conocimientos y habilidades que necesitaba. Estoy especialmente agradecido por haber aprendido tanto a través de las tareas.

Finalmente, estoy profundamente agradecido a mi familia por su amor, apoyo y ánimo incondicionales. Mis padres siempre han estado ahí para apoyarme tanto emocional como económicamente, y han hecho muchos sacrificios a lo largo de mi educación. Mi hermana y mi hermano nunca me han dejado solo durante mi proceso educativo y siempre han estado dispuestos a escucharme y darme consejos. Y a mi esposo, que ha sido mi compañero de vida y aprendizaje durante más de ocho años, le estoy inmensamente agradecido por su amor, apoyo y confianza incondicional. Su presencia ha sido una gran fuerza durante mi educación de maestría y una fuente de alegría en mi vida. El apoyo de cada uno de ellos ha sido invaluable para mí.





## INDICE DE CONTENIDOS

1 Introducción.....	1
1.1 Motivación.....	6
1.2 Objetivos.....	6
1.3 Organización de la memoria.....	6
2 Estado del Arte .....	9
2.1 Enfoques Existentes para la Detección de Malnutrición .....	9
2.1.1 Enfoques Existentes para la Detección de Malnutrición en Niños.....	9
2.1.2 Aprendizaje Automático (ML) para la Predicción de Malnutrición.....	9
2.2 Aprendizaje Automático en Spark.....	10
3 Diseño.....	11
3.1 Adquisición de Datos y Preprocesamiento.....	11
3.1.1 Descripción de Datos.....	11
3.1.2 Preprocesamiento y Análisis de Datos .....	12
4 Pruebas y Resultados .....	17
5 Conclusiones y Trabajo Futuro .....	23
6 Anexos .....	24
6.1 Anexo 1 – Distribución de Los Datos en Las Variables .....	24
6.2 Anexo 2 – Distribución de Clases en Target .....	27
6.3 Anexo 3 – Estadísticas de Variables Continuas en El Conjunto de Datos .....	27
6.4 Anexo 4 – Características más importantes en las fases de estudio del AA .....	27
6.5 Anexo 5 – Ajuste de Hiperparámetros en Cada Fase .....	28
Referencias .....	31

## INDICE DE FIGURAS

<i>FIGURA 1-1: FOTOGRAFÍAS DE DIFERENTES PAÍSES QUE RESALTEN LA IMPORTANCIA DE LA NUTRICIÓN INFANTIL. [3]</i> .....	1
<i>FIGURA 1-2: MEDIDAS ANTROPOMÉTRICAS EN NIÑOS. [3]</i> .....	4
<i>FIGURA 1-3: MALNUTRICIÓN INFANTIL. A) SOBREPESO INFANTIL. B) DEBILITANTE INFANTIL. C) RETRASO DEL CRECIMIENTO INFANTIL. [3], [42]</i> .....	5
<i>FIGURA 3-1: PAÍSES EN LOS CONJUNTOS DE DATOS. A) PAÍSES INCLUIDOS EN EL CONJUNTO DE DATOS DE LA BASE DE DATOS GLOBAL DE MALNUTRICIÓN INFANTIL DE LA OMS. B) PAÍSES INCLUIDOS EN EL CONJUNTO DE DATOS DE CHILDREN’S RECODE DEL PROGRAMA DHS.</i> .....	12
<i>FIGURA 3-2: CORRELACIONES ENTRE LAS PROPIAS CARACTERÍSTICAS</i> .....	14
<i>FIGURA 3-3: CORRELACIONES ENTRE LAS CARACTERÍSTICAS Y LA VARIABLE OBJETIVO</i> .....	15
<i>FIGURA 4-1: UNA MATRIZ DE CONFUSIÓN DE UN MODELO DE BOSQUE ALEATORIO PARA EVALUAR EL RENDIMIENTO DE UN MODELO DE CLASIFICACIÓN MULTICLASE. LOS ELEMENTOS DIAGONALES MUESTRAN EL NÚMERO DE PREDICCIONES CORRECTAS, MIENTRAS QUE LOS ELEMENTOS FUERA DE LA DIAGONAL INDICAN LAS INCORRECTAS.</i> .....	18

<i>FIGURA 4-2: MATRIZ DE CONFUSIÓN DEL BOSQUE ALEATORIO PARA CLASIFICACIÓN BINARIA .....</i>	<i>19</i>
--	-----------

<i>FIGURA 4-3: EVALUACIÓN DE TODAS LAS FASES EN UN GRÁFICO DE RADAR: A) CUATRO EVALUACIONES EN LA FASE I PARA CLASIFICACIÓN MULTICLASE B) CINCO EVALUACIONES EN LA FASE 2 PARA CLASIFICACIÓN BINARIA C) CINCO EVALUACIONES EN LA FASE 3 PARA CLASIFICACIÓN BINARIA DESPUÉS DE ELIMINAR LAS ETIQUETAS DE MALNUTRICIÓN, BAJO PESO Y SOBREPESO. ....</i>	<i>20</i>
---	-----------

## INDICE DE TABLAS

<i>TABLA 2-1: ESTUDIOS SIMILARES UTILIZANDO EL CONJUNTO DE DATOS DEL PROGRAMA DHS [6], [9], [24], [33], [34], [43], [47]. ....</i>	<i>10</i>
--	-----------

<i>TABLA 4-1: EVALUACIÓN DE LOS MODELOS DE APRENDIZAJE AUTOMÁTICO DESPUÉS DEL ENTRENAMIENTO EN LA FASE I CON DIFERENTES TÉCNICAS .....</i>	<i>18</i>
--	-----------

<i>TABLA 4-2: EVALUACIÓN DE LOS MODELOS DE APRENDIZAJE AUTOMÁTICO DESPUÉS DEL ENTRENAMIENTO EN LA FASE II CON DIFERENTES TÉCNICAS.....</i>	<i>19</i>
--	-----------

<i>TABLA 4-3: EVALUACIÓN DE LOS MODELOS DE APRENDIZAJE AUTOMÁTICO DESPUÉS DEL ENTRENAMIENTO EN LA FASE III CON DIFERENTES TÉCNICAS .....</i>	<i>19</i>
--	-----------

# 1 Introducción

---

La nutrición es fundamental para la salud de las personas, especialmente para el desarrollo de un niño. Una nutrición equilibrada o adecuada puede considerarse la obtención suficiente de nutrientes esenciales para mantener el equilibrio corporal y favorece un sistema inmunológico más fuerte, reduce el riesgo de enfermedades y promueve una larga vida [1], [2]. En la actualidad, la nutrición adecuada de los niños es de gran importancia para que puedan aprender mejor, ser más productivos y prevenir enfermedades, es decir, para llevar una vida saludable [1], [2].



**Figura 1-1:** Fotografías de diferentes países que resalten la importancia de la nutrición infantil. [3]

La nutrición insuficiente y desequilibrada son problemas de salud pública a nivel mundial [4], [5], [6]. Se caracterizan por una deficiencia, exceso o desbalance en el consumo de nutrientes esenciales [7], [8], [9]. En otras palabras, es la carencia de los nutrientes básicos para cubrir las necesidades de crecimiento [10]. Se ha observado que los niños que sufren de desnutrición y alimentación desequilibrada presentan problemas graves como un mayor

riesgo de enfermedades crónicas o intestinales, disfunciones del sistema inmunológico, alta mortalidad y dificultades en su desarrollo físico y cognitivo [1], [11], [12], [13], [14], [15], [16], [17], [18]. Una buena alimentación, especialmente en los primeros años de vida, es fundamental para el desarrollo cerebral, y los efectos negativos de la desnutrición pueden deteriorar las funciones cerebrales y reducir la productividad [5], [10]. Además, los problemas de nutrición en niños menores de cinco años pueden aumentar los gastos en salud en etapas posteriores de la vida [14], [15].

El consumo excesivo de alimentos en un niño (sobrealimentación) puede conducir a la obesidad en etapas posteriores de la vida y aumentar el riesgo de desarrollar diversas enfermedades como enfermedades cardíacas, accidentes cerebrovasculares y diabetes [9], [14], [19]. La obesidad se define como una acumulación excesiva de grasa en el cuerpo y se calcula mediante el índice de masa corporal (IMC), que se obtiene dividiendo el peso en kilogramos entre la altura en metros al cuadrado ( $\text{kg/m}^2$ ). Generalmente, un IMC inferior a 18.5 se considera bajo peso, entre 18.5 y 24.9 peso normal, entre 25 y 30 sobrepeso y superior a 30 obesidad [7], [20], [21]. Por ejemplo, en los Estados Unidos, el alto consumo de alimentos procesados, como la comida rápida, y su fácil acceso han llevado a una alimentación desequilibrada y, como consecuencia, a una mayor prevalencia de obesidad en comparación con la delgadez. Además, la disminución de la actividad física contribuye al aumento de la obesidad [14].

La desnutrición infantil implica una ingesta insuficiente de alimentos, lo que se traduce en una carencia de vitaminas y minerales esenciales. Esta situación puede conducir a problemas de crecimiento como wasting (bajo peso para la altura) y stunting (baja altura para la edad), e incluso a la muerte [14], [19]. Además de la desnutrición general, la deficiencia de micronutrientes es un problema común en la infancia. Las deficiencias más frecuentes son las de hierro, vitamina A, yodo y zinc [16], [22]. Si nos referimos a las enfermedades causadas por estas deficiencias, la deficiencia de hierro disminuye el rendimiento laboral y la resistencia, y aumenta el riesgo de muerte en niños. La deficiencia prolongada de vitamina A puede causar ceguera infantil (prevenible). La deficiencia de zinc aumenta el riesgo de neumonía, malaria y diarrea. La deficiencia de yodo tiene efectos negativos en los niños, aumentando el riesgo de muerte infantil [22]. Por ejemplo, en países africanos, debido a factores como la sequía y los conflictos, los niños tienen dificultades para acceder a alimentos y, como consecuencia, no obtienen las vitaminas y minerales necesarios, lo que resulta en altas tasas de niños con bajo peso y baja altura en comparación con países de altos ingresos. Otro ejemplo es Sudáfrica, donde los niños enfrentan un alto riesgo de desnutrición aguda [23].

En resumen, mientras que en los países de altos ingresos el acceso a alimentos y servicios de salud es fácil, en los países de bajos ingresos no ocurre lo mismo, y esta situación continúa siendo un problema global [14], [19]. Como consecuencia, los países de bajos ingresos presentan tasas más altas de malnutrición y sus problemas de salud asociados [6], [7], [24], [25]. En otras palabras, una buena nutrición está relacionado con el bienestar de los niños y, por extensión, de toda la sociedad [6], [18], [25]. Además del bajo ingreso, otros factores como enfermedades preexistentes, el nivel educativo de la madre, el número de hijos, los ingresos familiares y la calidad del agua potable también influyen en la desnutrición [5]. Asimismo, el tipo y la calidad de los alimentos consumidos, las diferencias entre zonas rurales y urbanas, y el acceso a servicios de salud son factores adicionales que pueden afectar la malnutrición [26]. Los países de Asia y África son los más afectados por la malnutrición

en comparación con otras regiones [19], [27]. En particular, la tasa de mortalidad infantil en el África subsahariana es 15 veces mayor que en los países de altos ingresos [28].

A nivel mundial, cuando se analizan niños menores de 5 años, la tasa de mortalidad supera los 10 millones al año [13]. La malnutrición es una de las principales causas de estas muertes, representando alrededor del 45% de los fallecimientos en este grupo de edad [2], [14], [18], [19], [25], [29], [30], [31]. Afortunadamente, la malnutrición es una condición tratable [32]. Las muertes infantiles relacionadas con la malnutrición pueden prevenirse a través de la vacunación, una atención y alimentación adecuadas, la lactancia materna y el acceso a servicios de salud y alimentos [28]. Estas medidas no solo reducen la malnutrición sino también los costos de salud [32]. Organizaciones internacionales como la OMS y UNICEF se centran en este tema y continúan sus estudios con encuestas en distintos países en determinados periodos, aunque no todos los años. Según sus datos, en 2019, 149 millones de niños eran de baja altura (stunted), 50 millones estaban demasiado bajo peso y 38,2 millones tenían sobrepeso [14]. La OMS estima que, en 2020, 149,2 millones de niños eran de baja altura y se ha fijado el objetivo de reducir esta cifra en un 50% para 2030 [27], [33], [34], [35]. Al mismo tiempo, 45,4 millones de niños estaban delgados (bajo peso para la altura) y 38,9 millones tenían sobrepeso, siendo las tasas de obesidad y sobrepeso más altas en los países de altos ingresos [2], [27]. Por su parte, UNICEF reportó en 2022 que había 148,1 millones de niños con baja altura, 45 millones con bajo peso y 37 millones con sobrepeso a nivel mundial [16], [36].

La OMS utiliza estándares específicos para recopilar datos cuantitativos y monitorear el crecimiento y desarrollo de niños en todo el mundo [35]. Mediante la medición antropométrica, que evalúa características físicas como la altura, el peso y la edad en niños de 0 a 5 años, la OMS establece estándares de crecimiento infantil. Estos estándares se representan en curvas de crecimiento para niños, permitiendo identificar casos de malnutrición a nivel mundial [35]. Gracias a estos estándares, la OMS puede implementar medidas para prevenir la malnutrición infantil. Históricamente, se utilizaron las referencias de crecimiento de NCHS (National Center for Health Statistics), una organización que recopila, analiza y difunde datos sanitarios sobre niños y adolescentes en los Estados Unidos de América, desde 1979, sin embargo, la OMS ha declarado que las curvas del NCHS no eran adecuados para monitorear el crecimiento infantil y que es necesario establecer nuevas normas [37]. Para ello, la OMS inició en 1997 el Estudio Multicéntrico de Referencia de Crecimiento (MGRS: Multicentre Growth Reference Study) para establecer nuevas curvas de crecimiento y ha publicado nuevas normas para sustituir a la Referencia Internacional de Crecimiento del NCHS/OMS (abril de 2006) [38], [39]. Para determinar los nuevos estándares, se consideraron niños que vivían en condiciones saludables y se utilizaron como referencia, también se tuvieron en cuenta características como la lactancia materna y el no fumar en las madres de estos niños [30], [38], [39].

Gracias a los resultados de estas investigaciones, la Organización Mundial de la Salud (OMS) recomienda que los bebés se alimenten exclusivamente con leche materna durante los primeros seis meses [10]. Además, en los nuevos estándares establecidos, la lactancia materna se considera el patrón de crecimiento normal para los bebés [37]. La alimentación exclusiva con leche materna durante los primeros 6 meses de vida es una de las estrategias más efectivas para prevenir la malnutrición y promover la salud a largo plazo [19]. La leche materna, además de ser un alimento seguro y óptimo para los bebés, proporciona anticuerpos protectores contra enfermedades y reduce el riesgo de obesidad infantil [40].



Asimismo, se ha demostrado que la lactancia materna reduce la mortalidad en niños menores de 5 años [22].

Si bien estos estándares de la OMS son ampliamente aceptados para abordar la malnutrición y la alimentación inadecuada en niños, para profundizar en la tesis, sería útil analizar los términos, fórmulas y técnicas de medición estadística que se utilizan en este contexto. En general, la malnutrición en los niños se manifiesta en forma de wasting (bajo peso para la altura), bajo peso (bajo peso para la edad), obesidad (sobrepeso para la altura) y retraso del crecimiento (baja altura para la edad) [17]. Para evaluar el estado nutricional de los niños, se utilizan comúnmente medidas antropométricas (como bajo peso, retraso del crecimiento, wasting (debilitante) y sobrepeso) [18], [29], [37].

**Medidas Antropométricas:** Dentro de las medidas antropométricas utilizadas para evaluar el crecimiento y desarrollo de los niños, se destacan con frecuencia el peso para la altura, la altura para la edad y el peso para la edad [11], [24], [29].



*Figura 1-2: Medidas antropométricas en niños. [3]*

**Peso para la altura (Wasting - Sobrepeso):** El wasting (debilitante) indica que un niño está delgado para su altura, lo cual sugiere una pérdida de peso significativa [7], [11], [17], [27]. Puede ser consecuencia de una malnutrición aguda, una enfermedad o una situación crónica adversa [11]. Aumenta el riesgo de mortalidad infantil, pero es reversible con un tratamiento adecuado [7]. Por otro lado, el sobrepeso indica que un niño pesa demasiado para su altura. Esto suele ocurrir cuando un niño consume más alimentos de los que necesita. A largo plazo, aumenta el riesgo de desarrollar enfermedades crónicas [7].

**Altura para la edad (Stunting):** El stunting (retraso del crecimiento) indica que un niño es demasiado bajo para su edad [17], [27], [34]. Suele ser causado por una alimentación inadecuada y condiciones socioeconómicas desfavorables. El stunting puede tener consecuencias a largo plazo, afectando a generaciones futuras [11], [27].

**Peso para la edad (Bajo peso):** El bajo peso indica que un niño pesa demasiado poco para su edad [7], [17], [27], [41].



**Figura 1-3:** Malnutrición infantil. a) Sobrepeso infantil. b) Debilitante infantil. c) Retraso del crecimiento infantil. [3], [42]

Los Estándares de Crecimiento Infantil de la OMS han establecido z-scores, obtenidos al restar el peso medio al peso observado y luego dividir la diferencia por la desviación estándar de la media, para las mediciones antropométricas, y estos valores se utilizan para determinar si un niño está creciendo adecuadamente. Los z-scores que caen por debajo de -2 o -3 desviaciones estándar se consideran significativamente bajos [11], [16], [43]. Por ejemplo, si el z-score de la talla para la edad está por debajo de -2 desviaciones estándar, se considera que el niño tiene baja estatura; si está por debajo de -3 desviaciones estándar, se considera que tiene talla extremadamente baja [30], [44]. La OMS define como "normal" un rango de z-scores entre -2 y 2 para la mediana de los estándares de crecimiento infantil. Esto significa que el niño se encuentra dentro del rango de peso y talla considerados saludables para su edad, ya que este rango corresponde a las desviaciones estándar en una población con distribución normal [11], [45], [46].

- **Debilitante (Wasting):** Peso para la altura < -2 desviaciones estándar de la media de los Estándares de Crecimiento Infantil de la OMS.
  - **Sobrepeso (Overweight):** Peso para la altura > +2 desviaciones estándar de la media de los Estándares de Crecimiento Infantil de la OMS.
  - **Retraso del crecimiento (Stunting):** Altura para la edad < -2 desviaciones estándar de la media de los Estándares de Crecimiento Infantil de la OMS.
  - **Bajo peso (Underweight):** Peso para la edad < -2 desviaciones estándar de la media de los Estándares de Crecimiento Infantil de la OMS.
- [24], [39], [47]



## **1.1 Motivación**

El Machine Learning es una rama de la inteligencia artificial que permite a las computadoras aprender de los datos y realizar tareas con datos no vistos previamente (datos de prueba). En otras palabras, esta rama permite entrenar a las computadoras para analizar datos, identificar patrones y correlaciones, y hacer predicciones basadas en esa información [23].

Hoy en día, podemos utilizar algoritmos de Machine Learning en diversas áreas como el diagnóstico de enfermedades, sistemas de recomendación, detección de fraudes y reconocimiento facial para la autenticación de identidad [6]. En nuestro conjunto de datos sobre malnutrición infantil, podemos utilizar estos algoritmos para predecir si un niño sufre de desnutrición, considerando factores como la región donde vive el niño o el estatus socioeconómico de su familia.

Estas predicciones pueden ayudarnos a encontrar soluciones más duraderas al problema de la malnutrición infantil. Por ejemplo, si la tasa de malnutrición infantil en Asia es mayor que en Europa, podemos desarrollar programas de alimentación específicos para niños en riesgo en estas regiones, optimizar la distribución de alimentos y costos, y fortalecer los servicios de salud para ofrecerles un futuro más saludable y erradicar así todos los tipos de malnutrición.

En los últimos años en el mundo, aunque limitados y escasos, el desarrollo de modelos que utilizan algoritmos de inteligencia artificial y Machine Learning y su uso en la predicción de condiciones como la malnutrición, el retraso del crecimiento, las enfermedades causadas por la malnutrición en niños y la mortalidad infantil sobre datos recogidos a través de encuestas han ido aumentando día a día [16]. Sin embargo, estos estudios se han limitado a determinados países o han utilizado un conjunto reducido de casos. Por lo tanto, es necesario realizar un estudio más exhaustivo sobre este tema, que incluya a diferentes países de forma conjunta.

## **1.2 Objetivos**

El objetivo de este estudio es utilizar algoritmos de Machine Learning para predecir la malnutrición infantil (retraso del crecimiento, debilitante, sobrepeso y bajo peso) utilizando características como la edad, el nivel educativo, el empleo, el estado civil de la madre, el nivel socioeconómico familiar, el número de hijos en la familia, el sexo y la edad del niño, y el tamaño al nacer. Los algoritmos de Machine Learning aprenderán los patrones y correlaciones ocultos entre estas características para proporcionarnos predicciones de alta precisión sobre la malnutrición infantil.

## **1.3 Organización de la memoria**

En el capítulo 2 se explicarán las mediciones y métodos utilizados para detectar la malnutrición, así como los resultados de algunos estudios sobre el tema. También se mencionarán estudios de aprendizaje automático sobre malnutrición. A continuación, se explicará por qué se utilizó Spark Aprendizaje Automático para este estudio.

El capítulo 3 proporcionará una breve descripción del conjunto de datos de la OMS y por qué no es del todo adecuado para el aprendizaje automático, aunque se puedan obtener

algunos resultados con él. También se discutirán las diferencias entre los conjuntos de datos del Programa DHS y de la OMS, el preprocesamiento realizado en el conjunto de datos Children's Recode, los países estudiados y las características utilizadas. Además, se describirán las correlaciones entre las características y las correlaciones de cada característica con el objetivo. La selección de características se realizó utilizando Random Forest después de entrenamiento.

El capítulo 4 explicará cómo se clasifica los datos, el número de fases de entrenamiento y prueba, el trabajo realizado en cada fase, los clasificadores utilizados y las tasas de éxito alcanzadas.

El capítulo 5 describirá las conclusiones extraídas y cómo se podrían mejorar futuras encuestas sobre malnutrición.

El capítulo 6 proporcionará las frecuencias de los valores en cada característica, el resumen estadístico de las características numéricas y qué características se utilizaron en cada fase.



## 2 Estado del Arte

---

### ***2.1 Enfoques Existentes para la Detección de Malnutrición***

#### **2.1.1 Enfoques Existentes para la Detección de Malnutrición en Niños**

Como se ha mencionado anteriormente, para determinar si un niño sufre de sobrepeso, retraso del crecimiento (stunting) o debilitante (wasting), se utilizan las medidas antropométricas y se comparan con los estándares de crecimiento infantil de la OMS. Estos estándares establecen valores de referencia y desviaciones estándar (Z-scores) que permiten identificar la presencia de malnutrición. Para realizar estas mediciones, se consideran la altura, el peso, la edad del niño e índice de masa corporal (IMC) para la edad. Además, se emplea el método de la Circunferencia Braquial Media (MUAC:Mid-Upper Arm Circumference), que consiste en medir el perímetro del brazo.

Si se pone ejemplos de diferentes estudios; un ejemplo de estudio utilizó datos antropométricos de niños de 0 a 59 meses de la Encuesta Demográfica y de Salud (DHS:Demographic and Health Surveys) para evaluar la calidad de estas mediciones a lo largo del tiempo. Aunque se observaron mejoras en la calidad de los datos, aún se identificaron diferencias significativas entre los estudios [47]. Además, un estudio comparó los estándares de crecimiento infantil de la OMS con las referencias de crecimiento de NCHS en estudios realizados en Ghana, India y Perú. Este estudio examinó la relación entre la lactancia materna y la malnutrición, así como si los indicadores de prevalencia de malnutrición en niños son predictores de mortalidad infantil [30]. Otro estudio utilizó la base de datos mundial de crecimiento infantil de la OMS para analizar datos de países de Asia, África, América Latina y Oceanía. Se encontró que la prevalencia de la malnutrición por proteínas y energía era más alta en los países asiáticos y más baja en los países latinoamericanos [43]. Un estudio adicional utilizó la base de datos mundial de crecimiento infantil de la OMS para evaluar las tendencias en la malnutrición a lo largo del tiempo. Al analizar las mediciones antropométricas, se observaron aumentos y disminuciones en la malnutrición entre diferentes continentes [18].

#### **2.1.2 Aprendizaje Automático (ML) para la Predicción de Malnutrición**

En cuanto a las predicciones realizadas utilizando algoritmos de Machine Learning, se han llevado a cabo diversos estudios. Por ejemplo, se seleccionaron países específicos como Papúa Nueva Guinea y se utilizaron datos de la Encuesta Demográfica y de Salud (DHS) de los años 2016-2018 para predecir el retraso en el crecimiento (stunting) en niños menores de 5 años. Además, se identificaron factores como la región, la edad, el nivel de bienestar familiar y el tamaño al nacer como influyentes en este problema [34]. De manera similar, en Etiopía, se emplearon datos de la DHS de 2016 para determinar la malnutrición y los factores sociodemográficos asociados [24]. En Bangladesh, se utilizaron datos de la Encuesta Demográfica y de Salud (DHS) de 2014 para predecir la malnutrición [6]. Otro estudio ha encontrado que la educación del hogar, el ingreso familiar, la localidad, el distrito y la edad son determinantes clave de la inseguridad alimentaria en niños de 6 meses a 5 años en Palestina [16]. En otras investigaciones, se ha predicho la malnutrición aguda severa en niños de 6 a 59 meses en el Hospital Jason Sendwe en Lubumbashi (República Democrática

del Congo) [25]. También se han explorado los efectos de la diarrea temprana en la infancia y la malnutrición en el desarrollo cognitivo posterior [13]. En otro estudio, se combinaron datos de la Base de Datos Global de la OMS sobre Crecimiento Infantil y Malnutrición y de los Informes sobre Desarrollo Mundial para analizar la relación entre los cambios en las tasas de mortalidad infantil y las modificaciones en el estado nutricional de los niños en países en desarrollo [12].

Para una perspectiva más general, se puede consultar la **Tabla 2-1** a continuación, que muestra el número de características, países, número de muestras, mejor modelo y acierto de algunos de los estudios de aprendizaje automático revisados. Estos estudios utilizaron etiquetas binarias y predijeron solo uno de los niños con retraso del crecimiento, debilitante y bajo peso. Aunque cada estudio utilizó características diferentes, algunos compartieron características comunes, como la educación de la madre, el tamaño del niño al nacer o el nivel económico.

	Núm. Caract.	País	Núm. Muestra	Mejor Modelo	Mejor Acierto
Shahriar et al. 2019	16	Bangladesh	6995	Multilayer Perceptron	84.93 %
Talukder et al. 2020	7	Bangladesh	6863	Random Forest	68.51 %
Hemo et al. 2021	13	Bangladesh	5000	Random Forest	72.40 %
Bitew et al. 2021	37	Etiopía	9471	Random Forest	88.10 %
Chilyabanyama et al. 2022	13	Zambia	6799	Random Forest	61.62 %
Shen et al. 2023	34	Papúa Nueva Guinea	3380	XGBoost	72.80 %
Ndagijimana et al. 2023	29	Ruanda	3814	Gradient Boosting	79.33 %

**Tabla 2-1:** Estudios similares utilizando el conjunto de datos del Programa DHS [6], [9], [24], [33], [34], [44], [48].

## 2.2 Aprendizaje Automático en Spark

Spark es una plataforma de procesamiento de datos sumamente útil e importante para gestionar grandes volúmenes de información. Para ejecutar tareas en Spark, podemos utilizar diferentes modos: modo local (para ejecutar en una máquina local), modo independiente o modo YARN (para conectarse a un planificador YARN).

En este estudio, empleamos Python como lenguaje de programación, utilizando PySpark que es un Python API para interactuar con las funcionalidades de Spark y sus algoritmos de aprendizaje automático para llevar a cabo nuestra investigación. Además, utilizamos las librerías de Python Matplotlib, Seaborn y Plotly para crear visualizaciones gráficas.

Una de las principales razones para utilizar PySpark es que estamos trabajando con un conjunto de datos grande. Los datos consisten en encuestas recopiladas de diferentes países. Dado que PySpark es un framework de computación distribuida, puede procesar cálculos en paralelo, lo que lo hace ideal para grandes conjuntos de datos. Para algoritmos de aprendizaje automático, Python también tiene las librerías scikit learn y tensorflow. Tensorflow puede trabajar con GPUs y TPUs que requieren configuración adicional para el procesamiento de grandes datos y es más adecuado para el aprendizaje profundo. Aunque scikit learn también puede trabajar con grandes conjuntos de datos, Pyspark tiene más ventajas para nosotros. Por ejemplo, PySpark tiene una mayor tolerancia a fallas en comparación con otras librerías, lo cual hace a PySpark especialmente útil para este estudio.

## 3 Diseño

---

### 3.1 Adquisición de Datos y Preprocesamiento

#### 3.1.1 Descripción de Datos

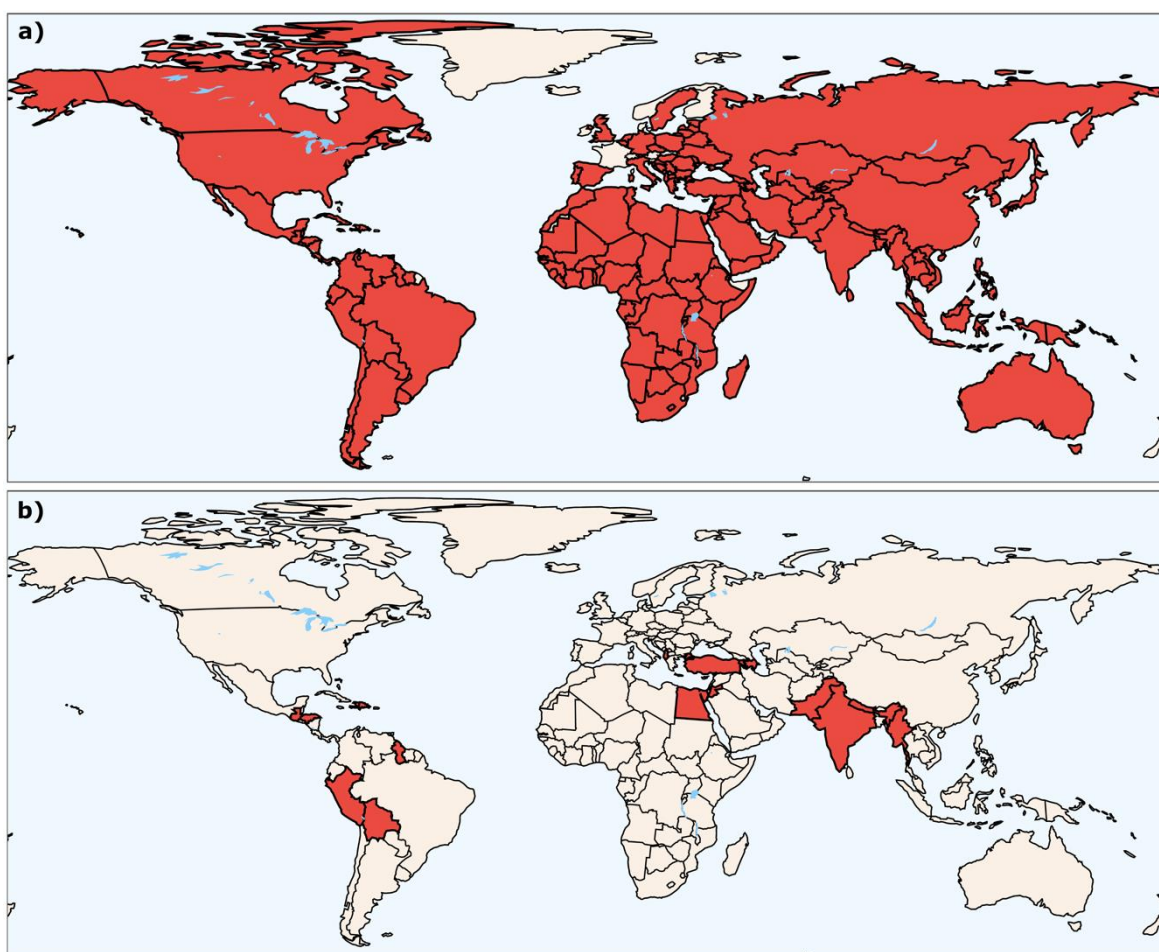
OMS ha comenzado a establecer estándares para monitorear la desnutrición infantil desde 1977 y, en 2006, mejoró los estándares y continúa el seguimiento obteniendo datos de todo el mundo. Si examinamos el conjunto de datos de la Base de Datos Global de Malnutrición Infantil de la OMS, es un conjunto de datos agregado que ha recopilado datos como resultado de encuestas realizadas en casi todos los países del mundo. La Base de Datos Global de Crecimiento y Malnutrición Infantil de la OMS se basa en encuestas poblacionales y debe cumplir ciertos criterios de calidad de datos [49]. Estos criterios incluyen ser basada en la población, incluir al menos 400 niños, realizar mediciones antropométricas, utilizar la población de referencia internacional NCHS/OMS y utilizar los puntos de corte de las puntuaciones  $z$  [29], [43]. En el conjunto de datos, se calcularon medidas antropométricas (peso para la altura, altura para la edad, peso para la edad e índice de masa corporal (IMC) para la edad) en niños de 0 a 5 años utilizando las puntuaciones  $z$  basadas en los estándares de crecimiento infantil de la OMS. Los niños que se encontraban por debajo de -2 o -3 desviaciones estándar se registró como porcentaje en las columnas Stunting, Wasting, Underweight y Overweight [49].

En estos datos agregados, se pueden procesar las columnas Urban/Rural, Educación de la madre y Quintil de riqueza por separado porque es un conjunto de datos agregado, pero no se pueden usar todas las columnas juntas. Si se analizan las columnas Urban/Rural, Stunting, Wasting, Underweight y Overweight, se observan que las áreas rurales tienen tasas de malnutrición más altas que las áreas urbanas. Cuando se consideran las columnas Educación de la madre, Stunting, Wasting, Underweight y Overweight, se pueden observar que a medida que aumenta la educación de la madre, disminuyen las tasas de malnutrición en los niños. Por otro lado, cuando se examinan las columnas Quintil de riqueza, Stunting, Wasting, Underweight y Overweight, se pueden observar que cuanto peor es el estado económico de la familia, mayor es la probabilidad de que los niños sufran malnutrición.

Dado que no se pueden procesar todas estas columnas juntas y el conjunto de datos de la OMS consiste en datos agregados, estos datos se vuelven inadecuado para el aprendizaje automático. Por lo tanto, para utilizar algoritmos de aprendizaje automático, necesitábamos acceso a los datos brutos. Por esta razón, accedimos a los datos de la encuesta del Programa de Encuestas Demográficas y de Salud (DHS) llamado Children's Recode. El Programa de Encuestas Demográficas y de Salud (DHS) ha realizado encuestas en más de 90 países, recopilando datos sobre población, salud y nutrición [50]. El conjunto de datos Children's Recode también incluye encuestas realizadas a niños de 0 a 5 años y sus madres en diversos países del mundo, principalmente en Asia, Europa, África y América Latina. Los datos antropométricos de los niños se presentan en forma de puntuaciones  $z$ , calculadas según los estándares de crecimiento infantil de la OMS.

### 3.1.2 Preprocesamiento y Análisis de Datos

Como se puede ver en la figura del mapa a continuación, mientras que hay casi 150 países en el conjunto de datos de la OMS, se pudo obtener 19 países de los conjuntos de datos de Children's Recode del Programa DHS. Dado que los datos brutos de la encuesta en el Programa DHS se registran por separado para cada país, para los datos de estos 19 países en Asia, Europa, África y América Latina, se combinaron los datos según estas características, teniendo en cuenta que las características que desea utilizar para los datos están presentes en todos los conjuntos de datos. Luego, durante la limpieza de los datos, se eliminaron los valores faltantes y también los valores inciertos como "Don't know" e "Inconsistent". Por lo tanto, los países que finalmente se obtuvieron y que se puede analizar son los siguientes: Azerbaiyán, Armenia, Albania, República Dominicana, Egipto, Bolivia, Honduras, Guatemala, India, Guyana, Haití, Myanmar, Nepal, Jordania, Maldivas, Perú, Pakistán, Turquía, Timor-Leste. Luego se comenzó a aplicar algoritmos de aprendizaje automático para estos países. Se utilizaron clasificadores como Gradient-boosted tree, Random Forest, Decision tree, Logistic Regression, Multilayer Perceptron, Linear Support Vector Machine, Naive Bayes, ajustando sus hiperparámetros. Los hiperparámetros ajustados se describen en el **Anexo 6.5**.



**Figura 3-1:** Países en los conjuntos de datos. **a)** Países incluidos en el conjunto de datos de la Base de Datos Global de Malnutrición Infantil de la OMS. **b)** Países incluidos en el conjunto de datos de Children's Recode del Programa DHS.

Después de fusionar los conjuntos de datos, se seleccionaron 20 características de entre 1293 características y el análisis se llevó a cabo sobre un conjunto de datos compuesto por 566.060 filas. Sin embargo, al iniciar el análisis, se eliminaron nuevamente algunas características y se trabajaron solo con las características más importantes. Se comparte las columnas del conjunto de datos con sus nombres de código para que, si alguien desea utilizarlas en el futuro, pueda comprender fácilmente los códigos en las columnas con sus nombres de código.

### 1. Columnas relacionadas con el niño

- **B4:** Sexo de los Niños
- **B8:** Edad de los Niños
- **B9:** Con Quién Vive de los Niños
- **B16:** Orden de Nacimiento (entre los niños nacidos)
- **V201:** Total de Niños Nacidos
- **V208:** Nacimientos en los Últimos Cinco Años
- **M18:** Tamaño de los Niños al Nacer

### 2. Columnas relacionadas con la madre

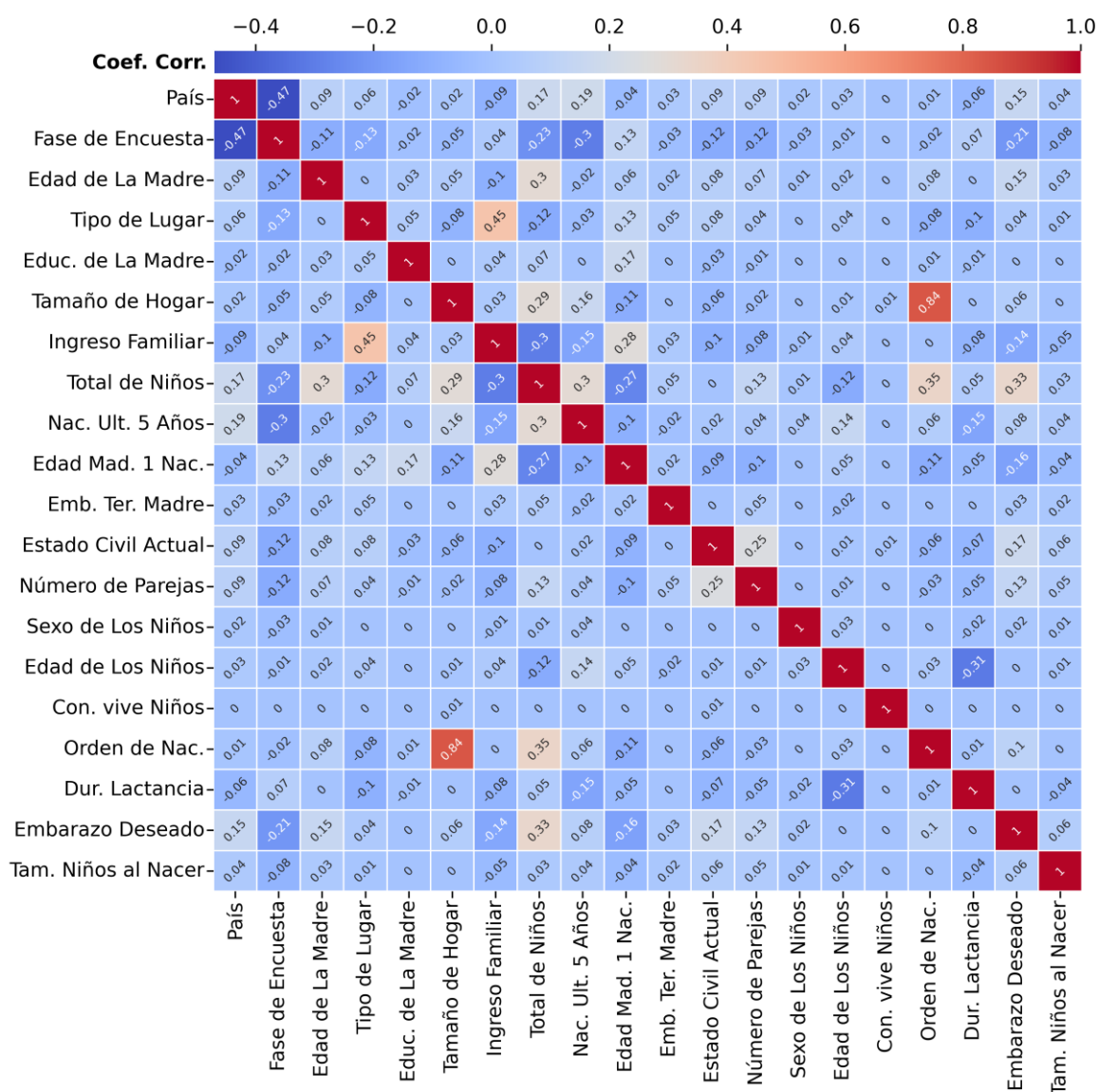
- **V013:** Edad por Grupo (de la madre)
- **V106:** Educación de la Madre
- **V212:** Edad al Primer Nacimiento
- **V228:** Alguna vez ha Terminado un Embarazo
- **V501:** Estado Civil Actual
- **V503:** Número de Parejas (de la madre)
- **M4:** Duración de la Lactancia Materna
- **M10:** Embarazo Deseado

### 3. Columnas relacionadas con el entorno en el que viven

- **V000:** Código de País (con Número de Fase)
- **V025:** Tipo de Lugar de Residencia
- **V136:** Tamaño de Hogar (Número de personas viviendo en un hogar)
- **V190:** Ingreso Familiar (Riqueza)

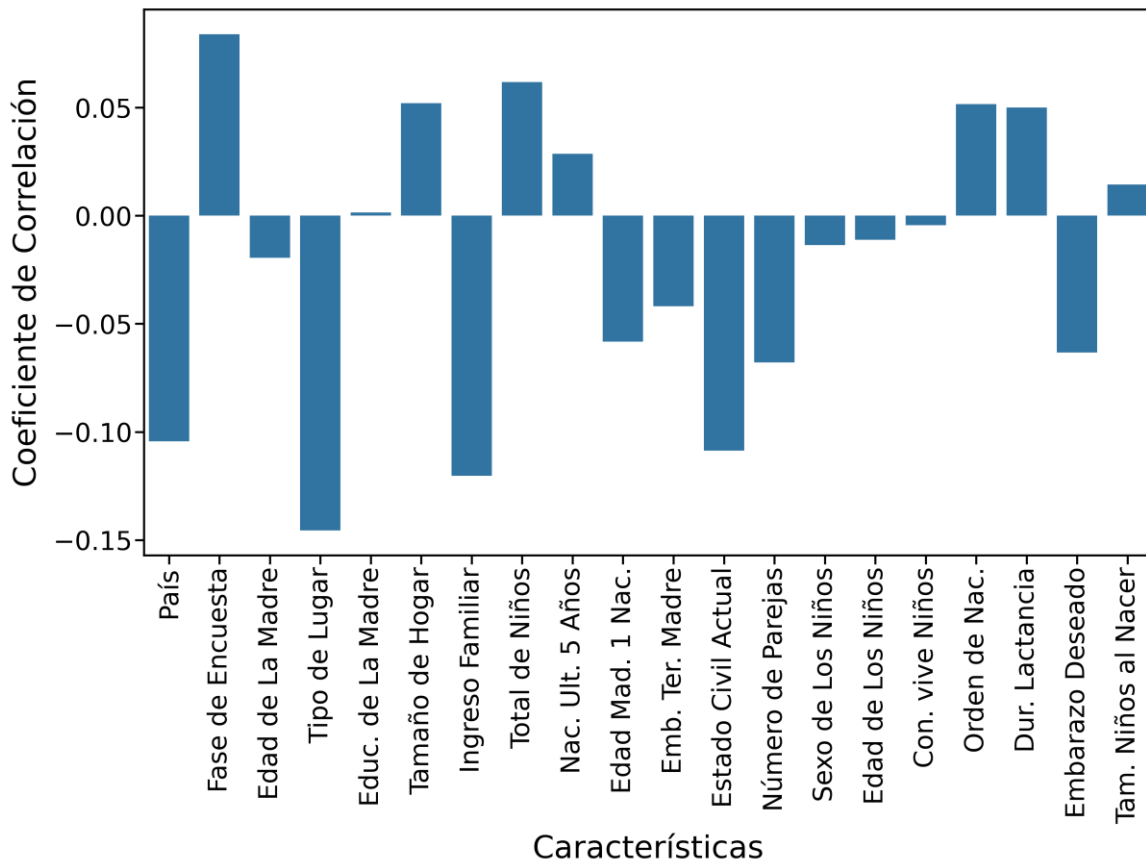
Se evaluó el estado nutricional de los niños utilizando las características de peso para la altura, altura para la edad y peso para la edad, siguiendo las normas de crecimiento de la OMS y se eliminaron estas características después de determinar etiquetas. Se clasificaron para los niños en las siguientes categorías: normal, retraso del crecimiento, debilitante, bajo peso y sobrepeso. Estas categorías fueron utilizadas como etiquetas objetivo en el modelo de aprendizaje automático. Como se muestra en la **Figura 3-2** y **Figura 3-3**, se analizaron tanto las correlaciones entre las propias características como las correlaciones entre las características y la variable objetivo una vez que todas las variables habían sido convertidas a un formato cuantitativo mediante la transformación de las variables categóricas en índices y luego en vectores.





**Figura 3-2:** Correlaciones entre las propias características

Al analizar la tabla de correlaciones, se encontró una fuerte correlación positiva (0.84) entre el tamaño de hogar y el orden de nacimiento. El número total de hijos presenta correlaciones positivas moderadas con varias variables, incluyendo embarazo deseado, orden de nacimiento, nacimientos en los últimos cinco años, tamaño de hogar y edad de la madre. Además, existe una correlación positiva moderada entre el ingreso familiar y el tipo de lugar. En términos de correlaciones negativas, se observaron relaciones inversas moderadas entre la fase de la encuesta (Fase: Versión de La Encuesta en El Programa DHS) y el número total de hijos, nacimientos en los últimos cinco años y embarazo deseado. La edad de la madre al primer nacimiento también tiene una correlación negativa moderada con el número total de hijos. Finalmente, existe una correlación negativa relativamente más fuerte entre el país y la fase de la encuesta.



**Figura 3-3:** Correlaciones entre las características y la variable objetivo

Considerando las correlaciones con la variable objetivo, se observan que país, tipo de lugar, ingreso familiar y estado civil actual exhiben una correlación inversa más fuerte con la variable objetivo que los demás. Por el contrario, tamaño de hogar, fase de encuesta, total de niños, orden de nacimiento y duración lactancia muestran una correlación positiva más débil, mientras que edad de madre en primero nacimiento, embarazo terminado de madre, número de parejas y embarazo deseado muestran una correlación inversa más débil.



## 4 Pruebas y Resultados

Después de limpiar el conjunto de datos e indexar las columnas categóricas, se entrenaron y probaron modelos de aprendizaje automático con las librerías de clasificación disponibles en PySpark. Hubo tres fases de entrenamiento para mejorar el acierto de los modelos de clasificación. La razón de esto se explicará en las siguientes partes a través de los resultados. Las tres fases se pueden enumerar como sigue:

1. Clasificación multiclase con el conjunto de datos completo
2. Clasificación binaria con el conjunto de datos completo
3. Clasificación binaria solo con datos de niños con retraso del crecimiento y normales

En todas las fases, todas las características en los conjuntos de datos que se consideraron correlacionadas con la malnutrición infantil se utilizaron directamente para el entrenamiento y luego se seleccionaron las características importantes mediante el bosque aleatorio (Random Forest), como se informa en el **Anexo 6.4**. Esta técnica de selección, en lugar de la correlación entre las características y el objetivo, se eligió porque el bosque aleatorio puede detectar relaciones no lineales. El entrenamiento y las pruebas se realizaron por separado con este conjunto de datos filtrado utilizando posiblemente diferentes modelos.

Las clasificaciones se realizaron con diferentes etiquetas de objetivo en cada fase. Los objetivos en la clasificación multiclase se determinaron como niños normales, con retraso del crecimiento, con debilitante, con bajo peso y con sobrepeso utilizando los estándares de la OMS para la malnutrición. Además, los objetivos en la clasificación binaria (en la segunda fase) son niños normales y malnutridos, etiquetando a los niños con retraso del crecimiento, con debilitante, con bajo peso y con sobrepeso como niños malnutridos. En la tercera fase, solo se eligieron niños normales y con retraso como objetivos. La siguiente tabla muestra los modelos utilizados en el estudio de ML.

Clasificación Multiclase	Clasificación Binaria
<ul style="list-style-type: none"><li>• Regresión Logística</li><li>• Árbol de Decisión</li><li>• Bosque Aleatorio</li><li>• Árbol de Gradiente Impulsado (Gradient Boosted Tree)</li><li>• Bayes Ingenuo (Naive Bayes)</li><li>• Perceptrón Multicapa</li></ul>	<ul style="list-style-type: none"><li>• Regresión Logística</li><li>• Árbol Decisión</li><li>• Bosque Aleatorio</li><li>• Árbol de Gradiente Impulsado (Gradient Boosted Tree)</li><li>• Máquina de Vectores de Soporte Lineal</li><li>• Bayes Ingenuo (Naive Bayes)</li><li>• Perceptrón Multicapa</li></ul>

### Fase I - Clasificación Multiclase

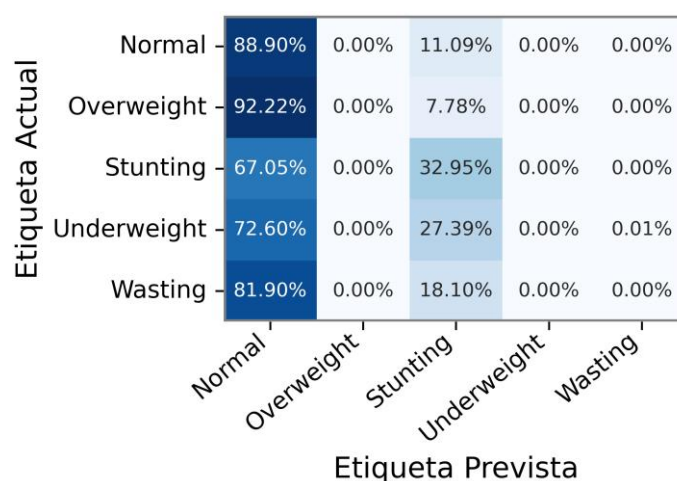
En la primera fase, se utilizó el conjunto de datos completo con 566.060 filas y, en primer lugar, se determinó la categoría a la que pertenecían los niños a partir de las columnas de peso para la altura, peso para la edad y altura para la edad. Luego, se realizó la afinación de hiperparámetros en estos modelos y utilizando una división de los datos en conjuntos de entrenamiento (80%) y prueba (20%), se evaluó el rendimiento de cada modelo. Los resultados se encontraron como se muestra en la **Tabla 4-1**. Es importante tener en cuenta

que no todos los modelos de ML son aplicables a la clasificación multiclase. Este análisis se centra en los modelos que son adecuados para tareas multiclase dentro de PySpark.

Modelos	Acierto (%)	Precisión (%)	Sensibilidad (%)	Puntuación F1 (%)
Regresión Logística	54.97	41.42	54.97	46.17
Árbol de Decisión	54.76	43.88	54.76	46.43
Bosque Aleatorio	54.97	41.42	54.97	46.15
Bayes Ingenuo	38.81	43.47	38.81	37.41
Perceptrón Multicapa	52.44	35.14	52.44	36.12

**Tabla 4-1:** Evaluación de los modelos de aprendizaje automático después del entrenamiento en la Fase I con diferentes técnicas

Estos valores demuestran que el éxito del entrenamiento es bajo. La matriz de confusión para el modelo de bosque aleatorio en la **Figura 4-1**, dada como ejemplo, muestra claramente el problema. En particular, tres etiquetas (niños con debilitante, bajo peso y sobrepeso) están funcionando mal y causan ruido en el entrenamiento. Aunque la clasificación multiclase es adecuada para problemas comunes en la vida diaria, los resultados muestran que el número de etiquetas diferentes no es suficiente para obtener un modelo exitoso. Por lo tanto, se requirió otra estrategia.



**Figura 4-1:** Una matriz de confusión de un modelo de bosque aleatorio para evaluar el rendimiento de un modelo de clasificación multiclase. Los elementos diagonales muestran el número de predicciones correctas, mientras que los elementos fuera de la diagonal indican las incorrectas.

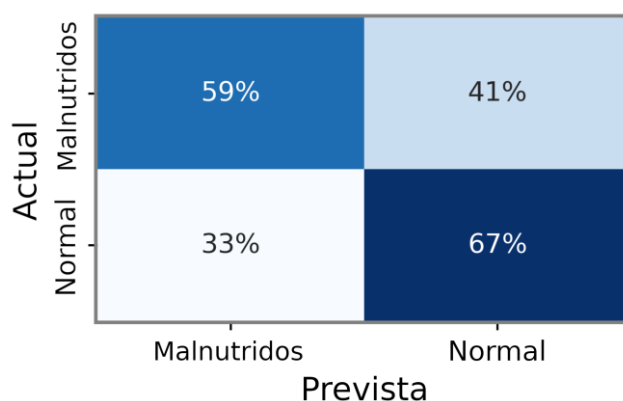
## Fase II - Clasificación Binaria

En esta fase, los niños con retraso del crecimiento, con debilitante, con bajo peso y con sobrepeso se etiquetaron como niños malnutridos, y el objetivo se convirtió en binario. Luego, se entrenaron los modelos de ML enumerados a continuación.

Estos valores son similares al resultado anterior. La matriz de confusión para el bosque aleatorio demuestra nuevamente el fracaso del modelo en la **Figura 4-2**. Diferentes etiquetas categorizadas bajo una etiqueta (niños malnutridos) causaron ruido y redujeron el éxito de los modelos. Por lo tanto, se necesitaba una estrategia diferente para un entrenamiento exitoso de ML.

Modelos	Acierto (%)	Precisión (%)	Sensibilidad (%)	Puntuación F1 (%)	AUC (%)
Regresión Logística	59.50	59.44	59.50	59.45	62.57
Árbol de Decisión	62.71	62.65	62.71	62.64	51.81
Bosque Aleatorio	63.22	63.16	63.22	63.16	68.22
Árbol de Gradiente Impulsado	63.17	63.11	63.17	63.09	68.29
Lineal SVM	58.66	58.88	58.66	58.68	62.16
Bayes Ingenuo	57.09	58.69	57.09	56.25	42.70
Perceptrón Multicapa	60.78	60.69	60.78	60.65	64.32

**Tabla 4-2:** Evaluación de los modelos de aprendizaje automático después del entrenamiento en la Fase II con diferentes técnicas



**Figura 4-2:** Matriz de confusión del bosque aleatorio para clasificación binaria

### Fase III - Clasificación Binaria con Etiquetas Específicas

En la última fase, se eliminaron las filas pertenecientes a niños con debilitante, con bajo peso y con sobrepeso, y los modelos de ML se entrenaron solo con datos de niños con retraso del crecimiento y normales. Este escenario crea una etiqueta binaria para que los mismos modelos de la fase anterior se puedan repetir como se muestra en la **Tabla 4-3** con evaluación.

Models	Acierto (%)	Precisión (%)	Sensibilidad (%)	Puntuación F1 (%)	AUC (%)
Regresión Logística	68.74	67.09	68.74	59.91	69.92
Árbol de Decisión	70.69	68.74	70.69	68.51	57.94
Bosque Aleatorio	70.80	68.76	70.80	67.85	72.23
Árbol de Gradiente Impulsado	70.76	68.78	70.76	68.44	72.34
Lineal SVM	67.44	45.48	67.44	54.32	63.19
Bayes Ingenuo	67.38	64.24	67.38	64.23	57.17
Perceptrón Multicapa	68.53	65.30	68.53	63.74	62.93

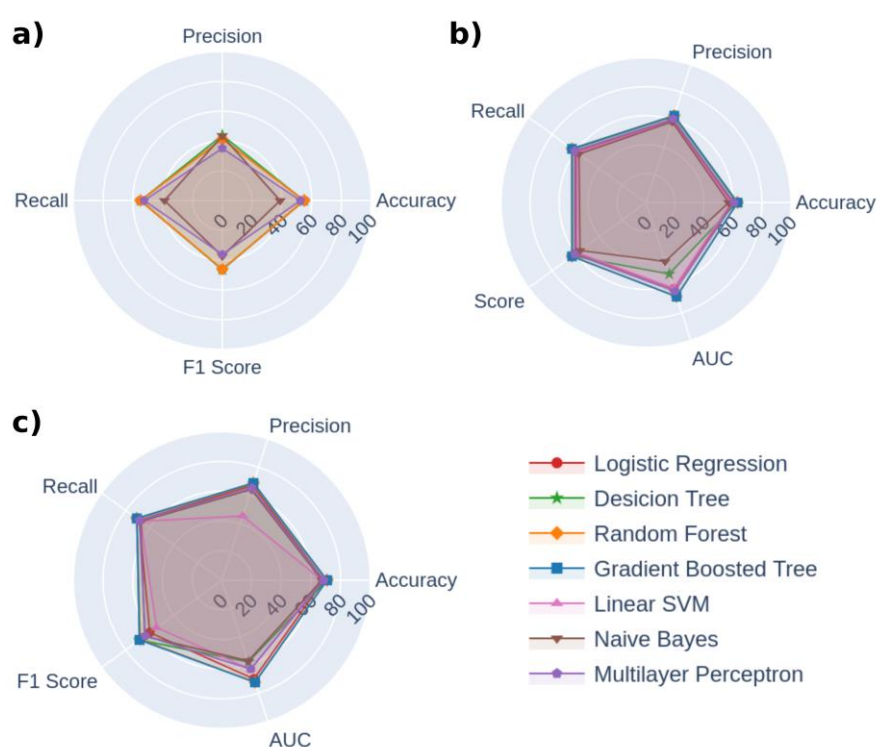
**Tabla 4-3:** Evaluación de los modelos de aprendizaje automático después del entrenamiento en la Fase III con diferentes técnicas

Aquí, los mejores resultados de los modelos de ML se obtuvieron con una precisión del 70.8%. En general, los resultados son más altos que en las fases anteriores como se muestra

en la **Figura 4-3**. Esto se debe a que usar objetivos simples en lugar de complejos e identificar/abordar el ruido dentro de los datos contribuye a mejorar los resultados.

El estudio y los resultados de ML arrojan muchas conclusiones importantes. Los modelos de ML en PySpark son adecuados para conjuntos de datos de la vida real, como la malnutrición infantil, pero requieren más muestras en cada etiqueta para entrenar los modelos. A pesar de emplear técnicas de sobremuestreo (oversampling) y submuestreo (undersampling) para equilibrar las clases en la variable objetivo, no se observó una mejora significativa en los resultados. Por lo tanto, tal estudio necesita más datos. Además, las encuestas con más preguntas completadas tienden a producir un mejor rendimiento del modelo, especialmente al comparar datos de diferentes países. Los análisis revelaron que las encuestas de diversas regiones a menudo tenían respuestas faltantes para preguntas específicas, lo que llevó a su exclusión del conjunto de datos. Esta reducción en el tamaño de la muestra afectó negativamente el éxito general de los modelos.

Las características, como la relevancia y la cantidad de preguntas, son tan cruciales como el tamaño de la muestra para lograr resultados exitosos del modelo. La OMS y el Programa DHS deberían considerar mejorar sus encuestas para incluir o reescribir preguntas más relevantes sobre la malnutrición infantil, mejorando así la efectividad de los modelos de ML.



**Figura 4-3:** Evaluación de todas las fases en un gráfico de radar: a) Cuatro evaluaciones en la Fase 1 para clasificación multiclase b) Cinco evaluaciones en la Fase 2 para clasificación binaria c) Cinco evaluaciones en la Fase 3 para clasificación binaria después de eliminar las etiquetas de malnutrición, bajo peso y sobrepeso.

Otra crítica que debe hacerse es sobre los países involucrados en este estudio. El conjunto de datos, compuesto principalmente por países con ingresos económicos bajos a medios-altos, carece de la diversidad necesaria para representar adecuadamente cada etiqueta

objetivo. Por ejemplo, si bien se puede predecir la falta de niños con sobrepeso en los países del conjunto de datos a partir de la situación económica de los países, no es útil para entrenar modelos de ML. Por lo tanto, es esencial recopilar datos de países con diferentes condiciones económicas para mejorar la efectividad del estudio.





## 5 Conclusiones y Trabajo Futuro

---

El objetivo de este estudio fue predecir la desnutrición en niños utilizando algoritmos de aprendizaje automático en Spark. Se utilizó el conjunto de datos Children's Recode del Programa DHS. Utilizando diferentes estrategias y aplicando diferentes modelos, la mayor precisión fue de 70.80% con el modelo Random Forest.

Como sugerencia para trabajos futuros, sería importante ampliar los estudios de encuesta a nivel mundial. En el conjunto de datos estudiado, los estudios se han realizado generalmente en países de bajos y medianos ingresos. Las encuestas realizadas en países de altos ingresos, así como en países de bajos y medianos ingresos, podrían proporcionar una imagen más definitiva de la detección y predicción de la desnutrición infantil. Además, será importante revisar las preguntas de la encuesta y hacer preguntas más específicas para la detección y predicción de la desnutrición. Además, en el futuro se pueden realizar estudios separados para multiclases y binarios a nivel mundial.

## 6 Anexos

---

### 6.1 Anexo 1 – Distribución de Los Datos en Las Variables

A continuación, se indican el número de variables y sus porcentajes en esa característica.

1. País

• IA (India)	: 366881 , 64.8 %
• PE (Perú)	: 52799 , 9.3 %
• EG (Egipto)	: 22978 , 4.1 %
• HN (Honduras)	: 18150 , 3.2 %
• JO (Jordania)	: 15032 , 2.7 %
• DR (República Dominicana)	: 13038 , 2.3 %
• TL (Timor-Leste)	: 11864 , 2.1 %
• GU (Guatemala)	: 11204 , 2.0 %
• NP (Nepal)	: 11466 , 2.0 %
• HT (Haití)	: 9029 , 1.6 %
• BO (Bolivia)	: 7103 , 1.3 %
• PK (Pakistán)	: 7132 , 1.3 %
• AL (Albania)	: 3719 , 0.7 %
• MM (Myanmar)	: 4017 , 0.7 %
• TR (Turquía)	: 3850 , 0.7 %
• AM (Armenia)	: 2790 , 0.5 %
• MV (Maldivas)	: 2275 , 0.4 %
• AZ (Azerbaiyán)	: 1422 , 0.3 %
• GY (Guyana)	: 1266 , 0.2 %

2. Fase (Versión de La Encuesta en El Programa DHS)

• F6	: 259511, 45.8 %
• F7	: 182845, 32.3 %
• F5	: 122118, 21.6 %
• F8	: 1541 , 0.3 %

3. V013 (Edad por Grupo de la Madre)

• 25-29	: 194899, 34.4 %
• 20-24	: 151835, 26.8 %
• 30-34	: 116407, 20.6 %
• 35-39	: 56439 , 10.0 %
• 15-19	: 20974 , 3.7 %
• 40-44	: 20335 , 3.6 %
• 45-49	: 5126 , 0.9 %

4. V025 (Tipo de Lugar de Residencia)

• Rural	: 387134, 68.4 %
• Urbano	: 178881, 31.6 %

5. V106 (Educación de la Madre)	
• Secundaria	: 253940, 44.9 %
• Sin educación	: 123602, 21.8 %
• Primaria	: 115811, 20.5 %
• Superior	: 72033 , 12.7 %
• Básica	: 629 , 0.1 %
6. V136 (Tamaño de Hogar) <sup>1</sup>	
• 2-5	: 267419, 47.2 %
• 6-7	: 160284, 28.3 %
• 8-10	: 97673 , 17.3 %
• 10+	: 40639 , 7.2 %
7. V190 (Ingreso Familiar)	
• Más pobre que todos	: 141455, 25.0 %
• Más pobre	: 129932, 23.0 %
• Medio	: 114732, 20.3 %
• Más rico	: 99099 , 17.5 %
• Más rico que todos	: 80797 , 14.3 %
8. V201 (Total de Niños Nacidos) <sup>1</sup>	
• 1-3	: 438594, 77.5 %
• 4-6	: 103342, 18.3 %
• 7+	: 24079 , 4.3 %
9. V208 (Nacimientos en los Últimos Cinco Años)	
• 1.0	: 350381, 61.9 %
• 2.0	: 185216, 32.7 %
• 3.0	: 28748 , 5.1 %
• 4.0	: 1588 , 0.3 %
• 5.0	: 75 , <0.1 %
• 6.0	: 7 , <0.1 %
10. V212 (Edad al Primer Nacimiento) <sup>1</sup>	
• 19-22	: 248298, 43.87 %
• 10-18	: 150659, 26.62 %
• 23-25	: 97062 , 17.15 %
• 25+	: 69996 , 12.37 %
11. V228 (Alguna vez ha Terminado un Embarazo)	
• No	: 468456, 82.76 %
• Sí	: 97559 , 17.24 %
12. V501 (Estado Civil Actual)	
• Casada	: 486699, 85.99 %

---

<sup>1</sup> Esta característica no es categórica, pero se agrupó para mostrar su distribución creando rangos significativos. La categorización no se utilizó en el estudio de Aprendizaje Automático.

• Conviviendo	: 46955 , 8.3 %
• Conviviendo con pareja	: 14289 , 2.52 %
• No conviviendo	: 6083 , 1.07 %
• Ya no conviviendo/separada	: 5035 , 0.89 %
• Viuda	: 3794 , 0.67 %
• Divorciada	: 3160 , 0.56 %
13. V503 (Número de Parejas de la Madre)	
• Una vez	: 539238, 95.27 %
• Más de una vez	: 26777 , 4.73 %
14. B4 (Sexo de Los Niños)	
• Niño	: 296893, 52.45 %
• Niña	: 269122, 47.55 %
15. B8 (Edad de Los Niños)	
• 1.0	: 128120, 22.64 %
• 0.0	: 123171, 21.76 %
• 2.0	: 114871, 20.29 %
• 3.0	: 104352, 18.44 %
• 4.0	: 95501 , 16.87 %
16. B9 (Con Quién Vive Los Niños)	
• Con su madre	: 565777, 99.96 %
• Vive en otro lugar	: 238 , 0.04 %
17. B16 (Orden de Nacimiento entre Los Niños Nacidos) <sup>1</sup>	
• 2-4	: 220779, 39.01 %
• 5-6	: 217261, 38.38 %
• 7-9	: 98819 , 17.46 %
• 10+	: 29156 , 5.15 %
18. M4 (Duración de la Lactancia Materna) <sup>2</sup>	
• Sigue amamantando	: 280989, 49.64 %
• 12-24	: 90738 , 16.03 %
• 6-12	: 49562 , 8.76 %
• Amamantó en el pasado, pero ya no lo hace	: 47293 , 8.36 %
• 0-6	: 40717 , 7.19 %
• Nunca amantó	: 30132 , 5.32 %
• 24+	: 26584 , 4.7 %
19. M10 (Embarazo Deseado)	
• Entonces	: 459367, 81.16 %
• Más tarde	: 55237 , 9.76 %

---

<sup>2</sup> Esta característica consta de variables categóricas y continuas. El programa DHS utiliza algunos números por las respuestas categóricas, mientras que otros números muestran la duración en meses. Se guardaron estos números para las etiquetas categóricas y se agruparon las duraciones para mostrar la distribución. En el aprendizaje automático, se utilizaron todos los números sin ninguna operación.

- Ya no : 51411 , 9.08 %
- 20. M18 (Tamaño de Los Niños al Nacer)
- Promedio : 371007, 65.55 %
- Superior al promedio : 86336 , 15.25 %
- Inferior al promedio : 64564 , 11.41 %
- Muy grande : 25636 , 4.53 %
- Muy pequeño : 18472 , 3.26 %

## 6.2 Anexo 2 – Distribución de Clases en Target

El número de clases en el objetivo con su proporción en total es el siguiente.

- Normal: 296844 (52.4 %)
- Retraso del crecimiento: 143323 (25.3 %)
- Debilitante: 52257 ( 9.2 %)
- Sobrepeso: 25926 ( 4.6 %)
- Bajo peso: 47710 ( 8.4 %)

## 6.3 Anexo 3 – Estadísticas de Variables Continuas en El Conjunto de Datos

Los resultados estadísticos de cada característica se presentan a continuación.

Característica	Núm. de Únicos	Mínimo	Máximo	Media	DE
V136	36	2	40	6.10	2.68
V201	18	1	18	2.78	1.82
V212	41	5	46	20.97	4.03
B16	31	2	35	5.26	2.12
M4	78	0	83	16.19	9.83

Las características V136, V201, V212, B16 y M4 en la tabla son el tamaño de hogar, el total de hijos nacidos, la edad de la madre al primer nacimiento, el orden de nacimiento de los niños y la duración de la lactancia materna, respectivamente.

## 6.4 Anexo 4 – Características más importantes en las fases de estudio del AA

Estas variables con códigos de programa DHS se consideraron las características más importantes para las diferentes fases de los estudios Aprendizaje Automático mediante Random Forest si sus valores de importancia eran superiores a 0,01:

- **Fase 1:**
  1. **País:** País en el que se ha realizado la encuesta. Se obtiene a partir del código V000 del programa DHS.
  2. **V025:** Tipo de Lugar de Residencia

3. **V106:** Educación de la Madre
  4. **V190:** Ingreso Familiar
  5. **V201:** Total de Niños Nacidos
  6. **V501:** Estado Civil Actual
  7. **B8:** Edad de Los Niños
  8. **M4:** Duración de la Lactancia Materna
- **Fase 2:**
    1. **País:** País en el que se ha realizado la encuesta. Se obtiene a partir del código V000 del programa DHS.
    2. **V025:** Tipo de Lugar de Residencia
    3. **V106:** Educación de la Madre
    4. **V190:** Ingreso Familiar
    5. **V201:** Total de Niños Nacidos
    6. **V212:** Edad al Primer Nacimiento
    7. **V501:** Estado Civil Actual
    8. **B8:** Edad de Los Niños
    9. **M18:** Tamaño de Los Niños al Nacer
  - **Fase 3:**
    1. **País:** País en el que se ha realizado la encuesta. Se obtiene a partir del código V000 del programa DHS.
    2. **V025:** Tipo de Lugar de Residencia
    3. **V106:** Educación de la Madre
    4. **V190:** Ingreso Familiar
    5. **V201:** Total de Niños Nacidos
    6. **B8:** Edad de Los Niños

## 6.5 Anexo 5 – Ajuste de Hiperparámetros en Cada Fase

La optimización de hiperparámetros con diferentes valores para los diferentes modelos indicados a continuación y los parámetros que dieron la mejor precisión se muestran en negrita. La optimización de hiperparámetros se llevó a cabo mediante ParamGridBuilder y CrossValidator, y se estableció el numFold en 3 para todos los modelos excepto el Decision Tree. La razón por la que estos métodos no se utilizan en el árbol de decisión es la fuga causada en la RAM, lo que hace que el sistema falle antes de encontrar los mejores parámetros. En lugar del método estándar, los siguientes parámetros para el Decision Tree se probaron uno por uno manualmente y se encontraron los mejores parámetros de esta manera.

- Phase 1
  - Logistic Regression
    - regParam: **0.01**, 0.1, 1.0
    - elasticNetParam: **0.0**, 0.5, 1.0
    - maxIter: **100**, 200, 300
  - Decision Tree
    - maxDepth: 5, **10**, 20

- maxBins: 16, **32**, 64
  - Random Forest
    - maxDepth: 5, **10**, 15
    - numTrees: 10, 20, **30**
  - Naive Bayes
    - smoothing: **0.0**, 0.5, 1, 2
  - Multilayer Perceptron
    - maxIter: 100, 200, **300**
    - stepSize: 0.03, 0.05, **0.1**
    - layers: [8, 7, 5], [8, 6, 5], [**8, 7, 6, 5**]
- Phase 2
  - Logistic Regression
    - regParam: **0.01**, 0.1, 1.0
    - elasticNetParam: **0.0**, 0.5, 1.0
    - maxIter: **100**, 200, 300
  - Decision Tree
    - maxDepth: 5, **10**, 20
    - maxBins: 16, 32, **64**
  - Random Forest
    - maxDepth: 5, 10, **15**
    - numTrees: 10, 20, **30**
  - Gradient Boosting
    - maxDepth: 3, 5, **7**
    - maxIter: 10, **20**, 30
  - Linear SVM
    - regParam: **0.01**, 0.1, 1
    - maxIter: 10, **100**, 1000
  - Naive Bayes
    - smoothing: 0.0, **0.5**, 1, 2
  - Multilayer Perceptron
    - maxIter: 100, 200, **300**
    - stepSize: **0.03**, 0.05, 0.1
    - layers: [**9, 5, 2**], [9, 7, 5, 2], [9, 6, 4, 2]
- Phase 3
  - Logistic Regression
    - regParam: 0.01, **0.1**, 1.0
    - elasticNetParam: **0.0**, 0.5, 1.0
    - maxIter: **100**, 200, 300
  - Decision Tree
    - maxDepth: 5, **10**, 20
    - maxBins: 16, **32**, 64
  - Random Forest
    - maxDepth: 5, **10**, 15



- numTrees: 10, 20, **30**
- Gradient Boosting
  - maxDepth: **3**, 5, 7
  - maxIter: 10, 20, **30**
- Linear SVM
  - regParam: **0.01**, 0.1, 1
  - maxIter: **10**, 100, 1000
- Naive Bayes
  - smoothing: **0.0**, 0.5, 1, 2
- Multilayer Perceptron
  - maxIter: 100, 200, **300**
  - stepSize: **0.03**, 0.05, 0.1
  - layers: **[6, 4, 2]**, [6, 4, 4, 2], [6, 4, 3, 2]

## Referencias

---

- [1] «Dietary Nutrition Cloud Platform Technology Based on Big Data», *IJMC*, vol. 2, n.º 1, mar. 2021, doi: 10.38007/IJMC.2021.020101.
- [2] «World Health Organization. Health Topic. Nutrition.» [En línea]. Disponible en: <https://www.who.int/health-topics/nutrition>
- [3] «WHO Global Photo Search. Malnutrition.» [En línea]. Disponible en: <https://whophotosearch.who.int/results?s%5Badv%5D=1&s%5Bkeywords%5D=malnutrition>
- [4] C. Shinsugi, D. Gunasekara, y H. Takimoto, «Use of Mid-Upper Arm Circumference (MUAC) to Predict Malnutrition among Sri Lankan Schoolchildren», *Nutrients*, vol. 12, n.º 1, p. 168, ene. 2020, doi: 10.3390/nu12010168.
- [5] J. M. Nagata, J. Gippet, S. Wager, A. Chavez, y P. H. Wise, «Prevalence and Predictors of Malnutrition among Guatemalan Children at 2 Years of Age», *PLoS ONE*, vol. 11, n.º 11, p. e0164772, nov. 2016, doi: 10.1371/journal.pone.0164772.
- [6] A. Talukder y B. Ahammed, «Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh», *Nutrition*, vol. 78, p. 110861, oct. 2020, doi: 10.1016/j.nut.2020.110861.
- [7] «World Health Organization. Malnutrition.» [En línea]. Disponible en: <https://www.who.int/en/news-room/fact-sheets/detail/malnutrition>
- [8] Z. Momand, P. Mongkolnam, P. Kositpanthavong, y J. H. Chan, «Data Mining Based Prediction of Malnutrition in Afghan Children», en *2020 12th International Conference on Knowledge and Smart Technology (KST)*, Pattaya, Chonburi, Thailand: IEEE, ene. 2020, pp. 12-17. doi: 10.1109/KST48564.2020.9059388.
- [9] M. M. Shahriar, M. S. Iqbal, S. Mitra, y A. K. Das, «A Deep Learning Approach to Predict Malnutrition Status of 0-59 Month's Older Children in Bangladesh», en *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, BALI, Indonesia: IEEE, jul. 2019, pp. 145-149. doi: 10.1109/ICIAICT.2019.8784823.
- [10] J. Hasegawa, Y. M. Ito, y T. Yamauchi, «Development of a screening tool to predict malnutrition among children under two years old in Zambia», *Global Health Action*, vol. 10, n.º 1, p. 1339981, ene. 2017, doi: 10.1080/16549716.2017.1339981.
- [11] M. de Onis y M. Blössner, «WHO Global Database on Child Growth and Malnutrition».
- [12] D. L. Pelletier y E. A. Frongillo, «Changes in Child Survival Are Strongly Associated with Changes in Malnutrition in Developing Countries», *The Journal of Nutrition*, vol. 133, n.º 1, pp. 107-119, ene. 2003, doi: 10.1093/jn/133.1.107.
- [13] P. D. Patrick *et al.*, «Early Childhood Diarrhea Predicts Cognitive Delays in Later Childhood Independently of Malnutrition», *The American Journal of Tropical Medicine and Hygiene*, vol. 95, n.º 5, pp. 1004-1010, nov. 2016, doi: 10.4269/ajtmh.16-0150.
- [14] T. Vassilakou, «Childhood Malnutrition: Time for Action», *Children*, vol. 8, n.º 2, p. 103, feb. 2021, doi: 10.3390/children8020103.
- [15] N. E. Lara-Pompa *et al.*, «Use of standardized body composition measurements and malnutrition screening tools to detect malnutrition risk and predict clinical outcomes in children with chronic conditions», *The American Journal of Clinical Nutrition*, vol. 112, n.º 6, pp. 1456-1467, dic. 2020, doi: 10.1093/ajcn/nqaa142.

- [16] R. Qasrawi *et al.*, «Machine Learning Approach for Predicting the Impact of Food Insecurity on Nutrient Consumption and Malnutrition in Children Aged 6 Months to 5 Years», *Children*, vol. 11, n.º 7, p. 810, jul. 2024, doi: 10.3390/children11070810.
- [17] M. Casapía, S. A. Joseph, C. Núñez, E. Rahme, y T. W. Gyorkos, «Parasite and maternal risk factors for malnutrition in preschool-age children in Belen, Peru using the new WHO Child Growth Standards», *Br J Nutr*, vol. 98, n.º 6, pp. 1259-1266, dic. 2007, doi: 10.1017/S0007114507795272.
- [18] M. de Onis y E. A. Frongillo, «Is malnutrition declining? An analysis of changes in levels of child malnutrition since 1980», *Bulletin of the World Health Organization*, 2000.
- [19] M. Sotiraki *et al.*, «Burden of Childhood Malnutrition: A Roadmap of Global and European Policies Promoting Healthy Nutrition for Infants and Young Children», *Children*, vol. 9, n.º 8, p. 1179, ago. 2022, doi: 10.3390/children9081179.
- [20] «WHO Obesity.» [En línea]. Disponible en: <https://www.who.int/health-topics/obesity>
- [21] «WHO Obesity and Overweight.» [En línea]. Disponible en: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>
- [22] T. Ahmed, M. Hossain, y K. I. Sanin, «Global Burden of Maternal and Child Undernutrition and Micronutrient Deficiencies», *Ann Nutr Metab*, vol. 61, n.º Supl. 1, pp. 8-17, 2012, doi: 10.1159/000345165.
- [23] S. Nel, U. D. Feucht, A. L. Nel, P. J. Becker, y F. A. M. Wenhold, «A novel screening tool to predict severe acute malnutrition through automated monitoring of weight-for-age growth curves», *Maternal & Child Nutrition*, vol. 18, n.º 3, p. e13364, jul. 2022, doi: 10.1111/mcn.13364.
- [24] F. H. Bitew, C. S. Sparks, y S. H. Nyarko, «Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia», *Public Health Nutr.*, pp. 1-12, oct. 2021, doi: 10.1017/S1368980021004262.
- [25] O. Mukuku *et al.*, «Predictive Model for the Risk of Severe Acute Malnutrition in Children», *Journal of Nutrition and Metabolism*, vol. 2019, pp. 1-7, jul. 2019, doi: 10.1155/2019/4740825.
- [26] A. Aziz Alimu y E. Prasetyo, «Predictors of Malnutrition in Children Aged Less than 5 Years in Surabaya, Indonesia», *Pakistan J. of Nutrition*, vol. 17, n.º 12, pp. 641-646, nov. 2018, doi: 10.3923/pjn.2018.641.646.
- [27] S. Jain, T. Khanam, A. J. Abedi, y A. A. Khan, «Efficient Machine Learning for Malnutrition Prediction among under-five children in India», en *2022 IEEE Delhi Section Conference (DELCON)*, New Delhi, India: IEEE, feb. 2022, pp. 1-10. doi: 10.1109/DELCON54057.2022.9753080.
- [28] «WHO Child Health.» [En línea]. Disponible en: <https://www.who.int/health-topics/child-health>
- [29] M. De Onis y M. Blössner, «The World Health Organization Global Database on Child Growth and Malnutrition: methodology and applications», *International Journal of Epidemiology*, vol. 32, n.º 4, pp. 518-526, ago. 2003, doi: 10.1093/ije/dyg099.
- [30] L. Vesel, B. Bahl, J. Martinez, N. Bhandari, y B. Kirkwood, «Use of the World Health Organization child growth standards to assess how infant malnutrition relates to breastfeeding and mortality», *Bull World Health Org*, vol. 88, n.º 1, pp. 39-48, ene. 2010, doi: 10.2471/BLT.08.057901.
- [31] «WHO Health Topic Malnutrition.» [En línea]. Disponible en: <https://www.who.int/health-topics/malnutrition>

- [32] B. T. Jin, M. H. Choi, M. F. Moyer, y D. A. Kim, «Predicting malnutrition from longitudinal patient trajectories with deep learning», *PLoS ONE*, vol. 17, n.º 7, p. e0271487, jul. 2022, doi: 10.1371/journal.pone.0271487.
- [33] S. Ndagijimana, I. H. Kabano, E. Masabo, y J. M. Ntaganda, «Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques», *J Prev Med Public Health*, vol. 56, n.º 1, pp. 41-49, ene. 2023, doi: 10.3961/jpmph.22.388.
- [34] H. Shen, H. Zhao, y Y. Jiang, «Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea», *Children*, vol. 10, n.º 10, p. 1638, sep. 2023, doi: 10.3390/children10101638.
- [35] «WHO Child Growth.» [En línea]. Disponible en: <https://www.who.int/health-topics/child-growth>
- [36] S. Wa, «Stunting affected an estimated 22.3 per cent or 148.1 million children under 5 globally in 2022».
- [37] K. Namakin, G. R. Sharifzadeh, M. Zardast, Z. Khoshmohabbat, y M. Saboori, «Comparison of the WHO Child Growth Standards with the NCHS for Estimation of Malnutrition in Birjand-Iran».
- [38] M. de Onis, «World Health Organization Reference Curves».
- [39] H. Yang y M. De Onis, «Algorithms for converting estimates of child malnutrition based on the NCHS reference into estimates based on the WHO Child Growth Standards», *BMC Pediatr*, vol. 8, n.º 1, p. 19, dic. 2008, doi: 10.1186/1471-2431-8-19.
- [40] «WHO Breastfeeding.» [En línea]. Disponible en: <https://www.who.int/health-topics/breastfeeding>
- [41] «UNICEF. Malnutrition.» [En línea]. Disponible en: <https://data.unicef.org/topic/nutrition/malnutrition/>
- [42] «The CSR Journal in India.» [En línea]. Disponible en: <https://thecsrjournal.in/child-obesity-rise-bhatia-hospital-remedial-measure/>
- [43] M. de Onfs, C. Monteiro, J. Akre, y G. Clugston, «The worldwide magnitude of protein-energy malnutrition: an overview from the WHO Global Database on Child Growth», vol. 71, 1993.
- [44] S. A. Hemo, y Md. I. Rayhan, «Classification tree and random forest model to predict under-five malnutrition in Bangladesh», *BBIJ*, vol. 10, n.º 3, pp. 116-123, sep. 2021, doi: 10.15406/bbij.2021.10.00337.
- [45] C. Browne *et al.*, «Multivariate random forest prediction of poverty and malnutrition prevalence», *PLoS ONE*, vol. 16, n.º 9, p. e0255519, sep. 2021, doi: 10.1371/journal.pone.0255519.
- [46] «WHO. Malnutrition in child.» [En línea]. Disponible en: <https://www.who.int/data/nutrition/nlis/info/malnutrition-in-children>
- [47] N. Perumal, S. Namaste, H. Qamar, A. Aimone, D. G. Bassani, y D. E. Roth, «Anthropometric data quality assessment in multisurvey studies of child growth», *The American Journal of Clinical Nutrition*, vol. 112, pp. 806S-815S, sep. 2020, doi: 10.1093/ajcn/nqaa162.
- [48] O. N. Chilyabanyama *et al.*, «Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia», *Children*, vol. 9, n.º 7, p. 1082, jul. 2022, doi: 10.3390/children9071082.
- [49] «WHO Data Platform.» [En línea]. Disponible en: <https://platform.who.int/nutrition/malnutrition-database>
- [50] «The Demographic and Health Surveys (DHS) Program.» [En línea]. Disponible en: <https://www.dhsprogram.com/>