



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

Análisis multivariante de la Encuesta Nacional sobre Desnutrición Infantil

PROYECTO INTEGRADOR

Previo a la obtención del Título de:

Matemático

Presentado por:

Jeremy Jardel Guadamud Pisco

GUAYAQUIL - ECUADOR

Año: 2024

DEDICATORIA

Este trabajo está dedicado a mi abuelita
Piedad, quién siempre quiso verme
graduado y sé que en el cielo está muy
feliz por mi logro.

Jeremy Guadamud P.

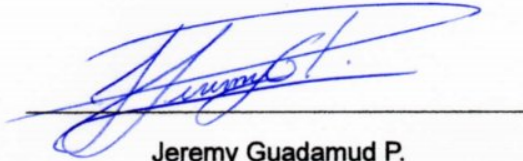
AGRADECIMIENTOS

Quiero expresar mi sincero agradecimiento. En primer lugar a Dios, ya que sin Él, no habría llegado hasta este punto en mi vida. También expreso mi agradecimiento a ESPOLE y a los profesores de la carrera por su dedicación y el valioso conocimiento impartido. No puedo dejar de agradecer enormemente a mi papá, mi mamá y mi hermano por el apoyo y la ayuda constantes durante mis estudios. Mis amigos Anthony, Javier, Aarón y Jhon han sido compañeros leales desde el preuniversitario, y a Esteban, Andrés y Carlos, matemáticos con los que compartí semestres y amanecidas de estudio, les agradezco sinceramente. Su apoyo ha sido fundamental en mi trayectoria académica.

Jeremy Guadamud P.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; *Jeremy Jardel Guadamud Pisco*, y doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



Jeremy Guadamud P.

EVALUADORES

Luz Elimar Marchan Mendoza

PROFESOR DE LA MATERIA

José Xavier Cabezas García

TUTOR DE PROYECTO

RESUMEN

La desnutrición infantil es una problemática y Ecuador no es la excepción. El objetivo de este proyecto es hacer un análisis multivariante de la Encuesta Nacional sobre Desnutrición Infantil 2022-2023 (ENDI) para detectar información relevante sobre distintos factores que afectan a la prevalencia de los distintos tipos de desnutrición en niños menores de cinco años. Es conocido que un análisis univariante o resultados descriptivos no son suficientes para comprender este gran problema. Se espera que este proyecto sea de ayuda para la toma de decisiones en busca de reducir la presencia de desnutrición infantil en el Ecuador. Se utilizó el software Rstudio para el análisis, depuración y preparación de las bases de datos. Luego, se implementaron los modelos Decision Tree y Random Forest para predecir, con base en varias variables, la presencia de desnutrición infantil, y de igual manera obtener las variables que más inciden en esa predicción. Después, a las variables más importantes se les aplicó la técnica MCA para detectar relaciones entre dichas variables. Se encontró que uno de los factores más determinantes es el lugar de origen del agua que beben los niños menores de cinco, así como factores demográficos y de condiciones de vida en los hogares de los niños. Finalmente, entre los niños menores de cinco años que sufren de desnutrición existen relaciones entre la contaminación que generan sus familias y los recursos que obtienen de la naturaleza.

Palabras Clave: Desnutrición Infantil, ENDI, Random Forest, MCA

ABSTRACT

Child malnutrition is a widespread issue, and Ecuador is no exception. The objective of this project is to make a multivariate analysis of the National Survey on Child Malnutrition 2022-2023 ENDI to identify relevant information regarding various factors that impact the prevalence of different types of malnutrition in children under the age of five. It is well known that a univariate analysis or descriptive results alone are insufficient to comprehend this significant problem. It is anticipated that this project will contribute to decision-making efforts aimed at reducing the prevalence of child malnutrition in Ecuador. Rstudio software was utilized for data analysis, cleaning, and preparation. Decision Tree and Random Forest models were then implemented to predict, based on various variables, the presence of child malnutrition and to identify the most influential factors in this prediction. Subsequently, the most important variables underwent Multiple Correspondence Analysis (MCA) to detect relationships among them. The study revealed that one of the most influential factors is the source of water consumed by children under the age of five. Additionally, demographic and living conditions in children's households play a significant role. Ultimately, among children under the age of five suffering from malnutrition, there are associations between their activities contributing to pollution generated by their families and the resources they obtain from nature.

Keywords: *Child malnutrition, ENDI, Random Forest, MCA.*

ÍNDICE GENERAL

RESUMEN	I
ABSTRACT	II
ABREVIATURAS	V
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	VIII
CAPÍTULO 1	1
1. INTRODUCCIÓN	1
1.1 Descripción del problema	1
1.2 Justificación del problema	3
1.3 Objetivos	4
1.3.1 Objetivo General	4
1.3.2 Objetivos Específicos	4
1.4 Marco teórico	5
1.4.1 Desnutrición Infantil	5
1.4.2 Desnutrición Infantil en el Ecuador	7
1.4.3 Encuesta Nacional de Desnutrición Infantil	8
1.4.4 Decision Tree y Random Forest	9
1.4.5 Análisis Multivariante	15
1.4.6 Análisis de Correspondencia Múltiples	16

CAPÍTULO 2	21
2. METODOLOGÍA	21
2.1 Tratamiento de datos	21
2.1.1 Análisis de las variables	24
2.2 Implementación de los modelos	26
CAPÍTULO 3	29
3. RESULTADOS Y ANÁLISIS	29
3.1 Resultados modelo Random Forest	29
3.1.1 Desnutrición Crónica	30
3.1.2 Desnutrición Aguda	31
3.1.3 Desnutrición Global	32
3.2 Resultados MCA	33
3.2.1 Desnutrición Crónica	33
3.2.2 Desnutrición Aguda	36
3.2.3 Desnutrición Global	39
CAPÍTULO 4	43
4. CONCLUSIONES Y RECOMENDACIONES	43
BIBLIOGRAFÍA	
APÉNDICES	

ABREVIATURAS

ESPOL	Escuela Superior Politécnica del Litoral
INEC	Instituto Nacional de Estadística y Censos
ENDEMAIN	Encuesta Demográfica y de Salud Materna e Infantil
ECV	Encuesta de Condiciones de Vida
ENSANUT	Encuesta Nacional de Salud y Nutrición
OMS	Organización Mundial de la Salud
ENDI	Encuesta Nacional sobre Desnutrición Infantil
UNICEF	Fondo de las Naciones Unidas para la Infancia
STECSDI	Secretaría Técnica Ecuador Crece Sin Desnutrición Infantil
DCI	Desnutrición Crónica Infantil
DAI	Desnutrición Aguda Infantil
DGI	Desnutrición Global Infantil
MCA	Análisis de Correspondencia Múltiple

ÍNDICE DE FIGURAS

Figura 1.1	Serie histórica de la DCI menores de 5 años	8
Figura 2.1	Árbol de decisión Desnutrición Crónica	27
Figura 2.2	Árbol de decisión Desnutrición Aguda	27
Figura 2.3	Árbol de decisión Desnutrición Global	28
Figura 3.1	Importancia de variables para determinar DC según Accuracy y Gini	30
Figura 3.2	Importancia de variables para determinar DA según Accuracy y Gini	31
Figura 3.3	Importancia de variables para determinar DG según Accuracy y Gini	33
Figura 3.4	Gráfico MCA con variables importantes para Accuracy en DCI	34
Figura 3.5	Contribución de las categorías de las variables para cada eje/dimensión	35
Figura 3.6	Gráfico MCA con variables importantes para Gini en DCI	35
Figura 3.7	Contribución de las categorías de las variables para cada eje/dimensión	36
Figura 3.8	Gráfico MCA con variables importantes para Accuracy en DAI	37
Figura 3.9	Contribución de las categorías de las variables para cada eje/dimensión	38
Figura 3.10	Gráfico MCA con variables importantes para Gini en DAI	38
Figura 3.11	Contribución de las categorías de las variables para cada eje/dimensión	39
Figura 3.12	Gráfico MCA con variables importantes para Accuracy en DGI	39
Figura 3.13	Contribución de las categorías de las variables para cada eje/dimensión	40
Figura 3.14	Gráfico MCA con variables importantes para Gini en DGI	41

Figura 3.15 Contribución de las categorías de las variables para cada eje/dimensión 42

ÍNDICE DE TABLAS

Tabla 1.1	Matriz de confusión	12
Tabla 1.2	Ejemplo: Matriz de confusión	13
Tabla 1.3	Niveles de kappa	14
Tabla 1.4	Ejemplo: Tabla de contingencia	18
Tabla A.1	Significado de las variables	

CAPÍTULO 1

1. INTRODUCCIÓN

En Ecuador, así como en el mundo está presente la desnutrición infantil de todo tipo, como lo pueden ser: crónica, aguda o global, y este es un problema muy serio para la salud del país. Según estimaciones de la Unicef (2021), Ecuador estaría ubicado como el cuarto país de América Latina y el Caribe con mayor presencia de desnutrición crónica infantil.

Frente a esta problemática surge este proyecto, con la intención de obtener y analizar información sobre la relación de la desnutrición infantil con diversos factores que hayan sido contemplados en la Encuesta Nacional de Desnutrición Infantil, la cual será la fuente de datos para el estudio. La finalidad de este proyecto radica en contribuir para una posible toma de decisiones en busca de una mejor salud para los niños y niñas ecuatorianos. Esto incentivado por evidencia de cómo este tipo de análisis puede ayudar a comprender la prevalencia de la desnutrición infantil en distintas regiones y por lo tanto, poder afrontar de manera eficiente la problemática.

1.1 Descripción del problema

La desnutrición infantil, caracterizada por la insuficiencia de nutrientes esenciales para un crecimiento y desarrollo saludable en los niños, ha sido una problemática que continúa representando un importante desafío de salud pública en Ecuador durante varios años. Esto se

refleja en los datos históricos del país obtenidos de la Encuesta Demográfica y de Salud Materna e Infantil (ENDEMAIN) 2004, la Encuesta de Condiciones de Vida (ECV) 2006 y 2014, la Encuesta Nacional de Salud y Nutrición (ENSANUT) 2012 y 2018; donde se ha registrado una prevalencia de desnutrición infantil en niños menores de 5 años que oscila entre el 28.9% y el 21.7% en el período de 2004 a 2018 (INEC, 2023b).

En este contexto, el 15 de diciembre de 2020 se implementó, mediante Decreto Ejecutivo 1211, la Estrategia Nacional Ecuador Crece Sin Desnutrición, que delegó al Instituto Nacional de Estadística y Censos (INEC) el diseño de una encuesta especializada sobre la desnutrición infantil. Ante ello, se crea la Encuesta Nacional de Desnutrición Infantil, conocida como ENDI, la cual se ha consolidado como una herramienta fundamental para evaluar la situación nutricional de los niños en el país (INEC, 2023a)

A pesar de los esfuerzos gubernamentales y programas de intervención que han contribuido a una disminución del 4.2%, los indicadores de desnutrición infantil continúan siendo motivo de preocupación. Según la Base de datos Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas, Joint Child Malnutrition Estimates (2021 Edition), Ecuador se sitúa como el cuarto país con la mayor prevalencia de desnutrición infantil en niños menores de cinco años, con un 17.5%. Esto lo ubica solamente por detrás de Honduras (19.9%), Haití (20.4%), y Guatemala (42.8%). Estas cifras resaltan la necesidad apremiante de abordar esta problemática de manera más efectiva.

Es conocido que la desnutrición tiene graves consecuencias en la salud, como el debilitamiento del sistema inmunológico, el retraso en el desarrollo físico y cognitivo, y una mayor susceptibilidad a enfermedades. La ENDI ha recopilado información de miles de niños

menores de cinco años y familias en todo el país, generando una base de datos rica en variables que incluye distintos aspectos que pueden ayudar a comprender la magnitud del problema y sus variaciones, si son analizados adecuadamente.

En este trabajo, se propone hacer un análisis multivariante de los datos obtenidos por la ENDI, los cuales están abiertos al público para la posible realización de estudios como este, centrado en la información recopilada sobre niños menores de cinco años; ya que la desnutrición infantil es un fenómeno de gran complejidad que no puede ser atribuido a un único factor aislado. El propósito es que ayude a identificar los factores de riesgo que contribuyen a la desnutrición y proporcione una base para la toma de decisiones y la implementación de políticas y programas para mejorar la salud y el bienestar de la población ecuatoriana.

Cabe recalcar que este estudio se enfoca en el conjunto del territorio ecuatoriano y se basa en la información recopilada durante la ENDI 2022-2023, otorgándole así relevancia y contemporaneidad a sus resultados.

1.2 Justificación del problema

La realización de un análisis multivariante en la Encuesta Nacional de Desnutrición Infantil (ENDI) 2022-2023 resulta esencial para comprender los factores clave que inciden en la desnutrición infantil en Ecuador. La desnutrición en niños es un asunto de enorme complejidad, sin ser atribuible a un único factor. Mediante un análisis multivariante, no solo se identifica sino también se cuantifica la influencia de múltiples factores en este problema, lo que proporciona una contribución significativa al campo de la salud. Se realiza con el propósito de aportar soluciones efectivas y políticas más fundamentadas para abordar este problema de manera integral.

La desnutrición infantil representa un desafío a nivel global, con implicaciones que abarcan

aspectos sociales, económicos y de salud pública. La comprensión de los factores determinantes no solo incide en el desarrollo individual de los niños, sino también en sus entornos y familias, con fuerte impacto en la sociedad en general y en el desarrollo sostenible del país.

Este proyecto busca contribuir a la formulación de estrategias que puedan ser implementadas a nivel gubernamental o comunitario para combatir la desnutrición infantil. Además, al entender los factores determinantes, se espera poder diseñar intervenciones más efectivas y específicas. El propósito de este proyecto es abordar un problema significativo y complejo que afecta directamente a la sociedad ecuatoriana.

1.3 Objetivos

1.3.1 Objetivo General

Detectar información relevante de los factores que contribuyen a la persistencia de la desnutrición infantil en el país mediante un análisis multivariante de la Encuesta Nacional de Desnutrición Infantil (ENDI) 2022-2023.

1.3.2 Objetivos Específicos

- Identificar grupos de variables que permitan realizar análisis multivariante;
- Utilizar métodos de reducción dimensional de variables que sirvan para resumir la información y que sean útiles en diferentes métodos multivariante;
- Analizar variables socioeconómicas, de condiciones de vida y de hábitos alimentarios más fuertemente relacionadas con la desnutrición infantil.

1.4 Marco teórico

En la presente sección, se describen fundamentos teóricos y técnicos relacionados con la problemática descrita anteriormente.

1.4.1 Desnutrición Infantil

Al hablar de desnutrición es importante saber que esta es un tipo de malnutrición, el término malnutrición hace referencia a carencias, excesos o desequilibrios en la cantidad de energía y/o nutrientes que una persona consume. Esta engloba dos principales grupos de condiciones. Uno de estos incluye a la obesidad, el sobrepeso y las enfermedades no transmisibles relacionadas con la dieta, como enfermedades cardíacas, cerebrovasculares, diabetes y cáncer (WHO, 2020). El otro grupo de malnutrición en la que se centra este proyecto es conocida como desnutrición, según ODS Territorio Ecuador (2018) es entendida como el resultado físico del hambre; debido a que incluye notables características físicas tales como retraso en el crecimiento, emaciación, peso bajo, y la falta o insuficiencia de micronutrientes (WHO, 2020).

Según la Organización Mundial de la Salud (OMS) la emaciación se caracteriza por tener un peso inferior al adecuado para la altura, indicando a menudo una posible pérdida de peso reciente y grave. En los niños, la desnutrición aguda o emaciación está vinculada a un mayor riesgo de mortalidad si no se aborda de manera apropiada. Por otro lado, el retraso en el crecimiento se define como una baja estatura para la edad, también conocida como desnutrición crónica (Palma, 2018); mientras que el bajo peso se refiere a tener un peso por debajo de lo normal para la edad. Un niño con bajo peso o desnutrición global (Palma, 2018) puede experimentar retraso en el crecimiento, emaciación o ambas en simultáneo. Se entiende por

deficiencia de micronutrientes a la carencia de vitaminas y minerales esenciales para distintas funciones corporales como la producción de enzimas, hormonas y otras sustancias necesarias para el crecimiento y desarrollo.

Ahora bien, la desnutrición infantil es una problemática presente alrededor de todo el mundo y a la que se ha prestado menos atención (Banco Mundial, 2021), pues el Director Ejecutivo de UNICEF Grant (1983) afirmó que, si los líderes políticos de las naciones caminaran juntos por una aldea o un pueblo en un país en desarrollo, sólo reconocerían alrededor del 2% de toda la desnutrición infantil a su alrededor.

El tamaño de este problema lo vemos reflejado en datos, pues en 2022 el retraso en el crecimiento afectó aproximadamente a 148.1 millones de niños menores de 5 años a nivel mundial, equivalente al 22.3% de infantes. De igual manera, la emaciación puso en peligro la vida de aproximadamente el 6.8% de niños menores de 5 años a nivel mundial, es decir, de 45 millones (UNICEF et al., 2023).

Por esta razón, a lo largo de los años, sea a nivel nacional, regional o global se han realizado distintos programas con el fin de llevar a cabo una recolección de datos, para poder conocer la situación de los países sobre esta temática. Esta iniciativa promueve el análisis de todo tipo a fin de obtener información que ayude a la disminución de la desnutrición infantil. Pues, como plantean UNICEF et al. (2023), "Todas las formas de malnutrición son prevenibles."

En México, se llevó a cabo un estudio donde se implementó algunos modelos logísticos multivariados entre desnutrición crónica y variables sociodemográficas. Obteniendo que la desnutrición crónica se asocia positivamente con habitar un hogar indígena y un factor protector de esta es la alta diversidad dietética (Cuevas-Nasu et al., 2019). En Cuba se realizó un estudio

observacional analítico a 148 niños de varios consultorios, en este caso los resultados permitieron conocer, como indica en su resumen, que “hábitos dietéticos inadecuados, la deficiente cultura sanitaria, el bajo per cápita, la disfunción familiar, la desvinculación laboral, el inadecuado desempeño en el cuidado de los hijos, la familia no nuclear y la baja escolaridad, estuvieron significativamente asociados a la desnutrición infantil” (Sánchez Hidalgo et al., 2020). Casos como estos muestran la importancia de analizar información sobre desnutrición infantil.

1.4.2 Desnutrición Infantil en el Ecuador

El Ecuador no es una excepción en la prevalencia de la desnutrición infantil, y por ende, en el intento de reducir la prevalencia de esta problemática. Se han llevado a cabo siete operaciones estadísticas para conocer de forma cuantitativa la situación alimentaria, nutricional y de salud de niños menores de cinco años desde 1986. (INEC, 2023b) Específicamente, estas operaciones fueron: el Diagnóstico de la Situación Alimentaria, Nutricional y de Salud de la Población Ecuatoriana Menor de Cinco Años en 1986; la Encuesta Demográfica y de Salud Materna e Infantil, dirigida por CEPAR en 2004; la Encuesta de Condiciones de Vida, dirigida por INEC en 2006; la Encuesta Nacional de Salud y Nutrición, llevado a cabo en conjunto por MSP e INEC en 2012; la Encuesta de Condiciones de Vida en 2014; la Encuesta Nacional de Salud y Nutrición, efectuada por INEC en 2018 y la Encuesta Nacional de Desnutrición Infantil en 2022-2023 siendo esta, la primera encuesta longitudinal especializada para estudiar la desnutrición en niños menores de 5 años y sus determinantes (INEC, 2023b). Estos estudios han permitido observar una mejora en el nivel de desnutrición a lo largo de la historia del Ecuador (ver Figura 1.1), teniendo concordancia con lo que nos plantea FAO et al. (2020), que en América Latina y el Caribe descende el retraso de crecimiento infantil, ya que la desnutrición

crónica se redujo de 22,7% en 1990 a 9% en 2019, un porcentaje inferior al promedio mundial de 21,3%.

Figura 1.1.
Serie histórica de la DCI menores de 5 años



Fuente: ENDEMAIN 2004, ECV 2006 y 2014, ENSANUT 2012 y 2018; Instituto Nacional de Estadística y Censos
* Referirse a la Nota Técnica para la comparabilidad de las cifras de desnutrición y sus informes anexos en: https://www.ecuadorencifras.gob.ec/encuesta_nacional_desnutricion_infantil/

INEC (2023b)

Pero, aún así, queda mucho camino por recorrer pues en Ecuador, entre 1993 y 2021 la curva de la DCI en menores de cinco años casi no se redujo (Unicef Ecuador, 2021). Incluso, según la Base de datos ODS de Naciones Unidas, Joint Child Malnutrition Estimates (2021 Edition), hecha por United Nations Children's Fund (UNICEF), World Health Organisation (WHO) y World Bank Group, Ecuador es el cuarto país con mayor prevalencia de desnutrición crónica infantil en Latinoamérica y el Caribe, con un 17.5%; además, a falta de información sobre Venezuela en esta base de datos, Ecuador se ubicaría como el país con más porcentaje de presencia de desnutrición crónica infantil de Sudamérica.

1.4.3 Encuesta Nacional de Desnutrición Infantil

A raíz de esta notable problemática presente en el país, mediante el Decreto Ejecutivo 1211 del 15 de diciembre de 2020, la Estrategia Nacional 'Ecuador Crece Sin Desnutrición' establece la

obligatoriedad de contar con una medición y un reporte de índices de desnutrición crónica infantil anualmente, a través del INEC, como uno de los cuatro ejes de gestión en la lucha contra la desnutrición crónica infantil, así enuncia la Vicepresidencia de la República del Ecuador.

En 2021, la Secretaría Técnica Ecuador Crece Sin Desnutrición Infantil (STECSDI) elabora el 'Plan Estratégico Intersectorial para la Prevención y Reducción de la Desnutrición Crónica Infantil'. El plan establece las prioridades del Estado y las inclinaciones del gobierno en la creación de un modelo social sostenido. Política que asegure que cada niño tenga oportunidades nutricionales óptimas desde el nacimiento para evitar la desnutrición crónica (ENDI, 2023).

La ENDI, a diferencia de todos los estudios anteriores que se realizaron en el país, es la primera encuesta especializada la cual entrega datos sobre Desnutrición Infantil y de igual manera describe el entorno social de los niños y niñas del Ecuador. Además, cuenta con el acompañamiento técnico del Fondo de las Naciones Unidas para la Infancia Unicef, Banco Mundial, Cepal y OPS (INEC, 2023a).

Esto permitió, al momento de realizar un análisis sobre la calidad de mediciones, determinar según el estándar actual de la OMS y UNICEF 2019 que la calidad de las mediciones de talla es de calidad aceptable en todas las encuestas anteriores, mientras que se cataloga como mediciones de alta calidad en la ENDI 2022-2023, evidenciando un salto de calidad con respecto a los estudios anteriores llevados a cabo en el país.(ENDI, 2023)

1.4.4 Decision Tree y Random Forest

"Decision Tree", o árbol de decisiones en español, es un modelo de machine learning supervisado. Los modelos de machine learning se clasifican, según la definición de clases,

principalmente en aprendizaje supervisado y aprendizaje no supervisado. El aprendizaje supervisado, también puede ser denominado como de clasificación, pues las clases son conocidas y están bien definidas en el conjunto de datos proporcionado, y el aprendizaje se realiza utilizando estas etiquetas de clase. Por ejemplo, clasificar una variable de respuesta con base en las demás variables. Mientras que, en el aprendizaje no supervisado las clases o sus límites no son conocidos, por lo que las propias etiquetas de clase también se aprenden, y las clases se definen en función de esto, por tal razón también se lo conoce como de agrupamiento (Suthaharan, 2016)

Existen dos tipos de árboles de decisiones: los árboles de regresión, los cuales tienen una variable de respuesta continua, y los árboles de clasificación, que tienen una variable de respuesta categórica. En términos sencillos, en un árbol de decisión, se divide la población en subpoblaciones basándose en el resultado de la aplicación de una función discreta escogida según algunas medidas de división. (Kotsiantis, 2013)

Una medida de división de datos es conocida como la impureza de Gini. La abreviatura Gini significa "generalized inequality index" o traducido al español, índice de desigualdad generalizada (Suthaharan, 2016), esta medida expresa la probabilidad de que dos muestras seleccionadas al azar del conjunto de datos, pertenezcan a categorías diferentes en la variable de respuesta (Zhou, 2021). Es decir, en el contexto de un nodo en un árbol de decisión, el Gini cuantifica qué tan mezcladas están las categorías en ese nodo. La impureza de Gini está definida como:

$$\text{gini}(D) = 1 - \sum_i^c p(i)^2 \quad (1.1)$$

Donde c es el número de categorías en la variable de respuesta del modelo, y $p(i)$ es la probabilidad de tener una observación de la categoría i en un conjunto de datos D , siendo su

fórmula general:

$$p(i) = \frac{\text{Número de observaciones de la clase } i}{\text{Total de observaciones en el conjunto de datos } D}$$

Así pues, el Gini de cada variable es conocido como "Weighted Gini Score" o gini ponderado, este viene representado como:

$$\text{gini}_v(D) = \sum_j^k \frac{n_j}{N_D} \cdot \text{gini}(j) \quad (1.2)$$

Aquí, v representa una variable específica, k es el número de categorías de la variable v , N_D es el número total de registros de la variable v en el conjunto D y $\text{gini}(i)$ es el gini de la categoría j .

Con el uso de la fórmula 1.2 se determinan los ginis ponderados de cada variable, con el objetivo de escoger la de menor valor, pues como menciona Zhou (2021) cuanto menor sea el Gini, mayor será la pureza del conjunto de datos D .

En síntesis, un árbol de decisiones predice una variable de respuesta (con datos asignados), con base en las demás variables de un grupo de datos. Esto, buscando mejorar la homogeneidad de los datos con respecto a la variable de respuesta, es decir, que las categorías estén menos mezcladas. Una forma común de evaluar la calidad de las predicciones dadas por un árbol de decisión es a través de la Matriz de Confusión. Una Matriz de Confusión nos muestra, en caso de variables binarias, una tabla con los resultados de los aciertos y desaciertos del modelo.

En la tabla 1.2 se tiene VP = Verdaderos Positivos, VN = Verdaderos Negativos, FP = Falsos Positivos y FN = Falsos Negativos. Cabe mencionar que una matriz de confusión nos brinda cuatro medidas cualitativas diferentes: acuracidad, precisión, sensibilidad y especificidad.

Tabla 1.1.
Matriz de confusión

Valores reales	Predicción	
	SI	NO
SI	VP	FN
NO	FP	VN

La acuracidad mide el rendimiento del modelo en función de la proporcionalidad entre los falsos positivos y los verdaderos positivos; si esta es alta, significa que la clasificación de ambas clases es altamente precisa y los falsos negativos y falsos positivos son descartables (Suthaharan, 2016).

Se expresa de la siguiente manera:

$$\text{Accuracy} = \frac{1}{1 + \frac{FP+FN}{VP+VN}} \quad (1.3)$$

Por otro lado, la medida de precisión, según Suthaharan (2016) describe el desempeño del modelo basado en la proporcionalidad entre los falsos positivos y los verdaderos positivos; si la precisión es alta, entonces la clasificación de una de las categorías es precisamente alta con pocos falsos negativos. Esta se define de la siguiente manera:

$$\text{Precision} = \frac{1}{1 + \frac{FP}{VP}} \quad (1.4)$$

La sensibilidad y especificidad miden el rendimiento del modelo con base en la proporcionalidad entre los falsos negativos y los verdaderos positivos, y los verdaderos negativos y los falsos positivos, respectivamente. Si la sensibilidad es alta, entonces la clasificación de los verdaderos positivos es altamente sensible a los falsos positivos, de igual manera si la especificidad es alta significa que los verdaderos negativos son altos con

importantes falsos negativos. Se calculan de la siguiente manera:

$$\text{Sensitivity} = \frac{1}{1 + \frac{FN}{VP}} \quad (1.5)$$

$$\text{Specificity} = \frac{1}{1 + \frac{FP}{VN}} \quad (1.6)$$

De igual manera, existe una medición que permite determinar que tan confiable es la predicción del modelo. Esta es la puntuación kappa de Cohen, la cual es una medida de la concordancia entre los resultados de la predicción y la acuracidad del mismo (Reeve and Gottlieb, 2020). El estadístico kappa se obtiene de la siguiente manera:

$$\kappa = \frac{\text{Accuracy} - P_e}{1 - P_e} \quad (1.7)$$

Donde P_e representa la probabilidad de acuerdos esperados por azar. Por ejemplo, tome un modelo de una muestra con una precisión (entiendase por acuracidad) del 75% y unas predicciones, tal como muestra la tabla 1.2:

Tabla 1.2.

Ejemplo: Matriz de confusión

Valores reales	Predicción	
	SI	NO
SI	30	10
NO	5	15

Entonces $P_e = \frac{35}{60} \times \frac{40}{60} + \frac{25}{60} \times \frac{20}{60} = 0.52$, resultando:

$$\kappa = \frac{0.75 - 0.52}{1 - 0.52} = 0.47$$

Para concluir, kappa toma valores en el rango [-1,1], donde 1 indica una concordancia perfecta, 0 indica una coincidencia de resultados igual a la esperada por azar y valores negativos

indican una concordancia peor que la esperada por azar. Al valor de kappa se le asigna un nivel de confiabilidad según el intervalo en el que se encuentre el número, como lo indica la tabla 1.3

Tabla 1.3.
Niveles de kappa

kappa-score	Nivel de confiabilidad
< 0	poor
0.01 – 0.2	slight
0.21 – 0.4	fair
0.41 – 0.6	moderate
0.61 – 0.8	substantial
0.81 – 1	almost perfect

Ahora bien, se habló de un modelo muy útil, como lo es un árbol de decisión, pero ¿qué pasaría si en vez de usar un solo árbol, se le da un mayor rol usando 500? Aquí es donde aparece el modelo Random Forest. Random Forest es, al igual que el árbol de decisión, un modelo predictivo supervisado de machine learning, este modelo es uno de los que tiene mejor rendimiento para construir conjuntos. Obtiene su fortaleza de dos aspectos: el uso de submuestras aleatorias de los datos de entrenamiento y la aleatorización del algoritmo para aprender clasificadores de nivel base usando árboles de decisión Kotsiantis (2013).

Random Forest brinda una mejora en la medidas de precisión de un árbol de decisión, pues es una paralelización de la aplicación de árboles de decisión, al mismo tiempo para la clasificación, a distintos grupos de entrenamiento de un conjunto de datos.

Así pues, este modelo brinda una predicción sobre una variable de respuesta con base en las demás variables, y el peso que ejercen estas sobre las funciones de medición o división.

Por ejemplo, según la acuracidad o según el Gini, el grupo de variables con mayor incidencia al momento de la predicción se consideran las variables más importantes, pues estas afectan directamente y en mayor medida a la elección de una categoría en la variable de respuesta.

1.4.5 *Análisis Multivariante*

El Análisis Multivariante consiste en explorar, analizar y comprender las conexiones entre varias variables mediante el desarrollo de complejos modelos estadísticos que permiten distinguir la contribución independiente de cada una de ellas en el sistema de relaciones, y a su vez permite describir, explicar o anticipar los fenómenos que son objeto de interés para la investigación (Meneses, 2019). De una manera más sencilla, Hair (2010) plantea al análisis multivariante como una extensión del análisis bivariante y este, a su vez, como una extensión del análisis univariante.

Es importante tener presente que la clave del análisis multivariante no radica en que los investigadores pueden realizar este análisis debido a la disponibilidad de múltiples variables. Este tipo de análisis es particularmente útil para abordar problemas complejos porque puede modelar las múltiples relaciones entre las diferentes variables involucradas en una investigación a la vez. Este es su rasgo distintivo (Meneses, 2019).

Así como los modelos de machine learning, las técnicas de análisis multivariante se pueden clasificar en técnicas de análisis de relaciones de dependencia y de interdependencia. Las técnicas de análisis de dependencia buscan la existencia o ausencia de relaciones entre dos grupos de variables que deben ser clasificadas según el contexto de cada estudio como variables dependientes y variables independientes. El objetivo de este tipo de técnica es determinar si el conjunto establecido como variables independientes afecta al conjunto de

variables dependientes de manera grupal o individualmente (Aldás and Uriel, 2017).

Por otro lado, las técnicas de análisis de interdependencia, se utilizan cuando no es posible dividir las variables en dependientes e independientes entre ellas, y por lo tanto el objetivo es determinar cómo y por qué están correlacionadas las variables (Aldás and Uriel, 2017).

Este proyecto se centra en la aplicación de estas técnicas, principalmente en el análisis de correspondencias múltiples (MCA, por sus siglas en inglés). Pues, se busca conocer el tipo y grado de relación que existe entre las distintas variables presentes en la Encuesta Nacional de Desnutrición Infantil, para así poder analizar sus datos y obtener información relevante que pueda ayudar a una futura toma de decisiones.

1.4.6 *Análisis de Correspondencia Múltiples*

Para poder hablar sobre el MCA, primero debemos saber que este es una extensión del Análisis de correspondencia simple (CA, por sus siglas en inglés), ya que, en el MCA se trabaja con variables categóricas con distintos números de categorías, o como dice Meneses (2019), todas las categorías contribuyen a conformar los ejes de una reducción dimensional. El MCA se utiliza comúnmente para analizar datos obtenidos a través de encuestas, por tal razón es la técnica a usar en este proyecto.

Al utilizar MCA se puede estudiar los datos según los individuos, las variables y las categorías (Husson et al., 2011). En este proyecto se ignorará a los individuos, pues son registros identificados con variables identificadoras, pero anónimos; es más, con la utilización de MCA nos centramos en el estudio de las categorías pues estas representan tanto a las variables como a los individuos que seleccionan una categoría específica.

Para analizar la cercanía entre las categorías, es necesario iniciar calculando la distancia

entre ellas. Por lo tanto, a cada una de dos categorías, se les llamará k_0 y k_1 , se les asignará un grupo de individuos; así pues el número de individuos que tienen la categoría k_0 o la categoría k_1 se expresa como $I_{k_0 \neq k_1}$, mientras que el número de individuos que tiene una categoría k se expresa como I_k . La distancia entre dos categorías k_0 y k_1 se calcula como:

$$d_{k_0, k_1}^2 = C \frac{I_{k_0 \neq k_1}}{I_{k_0} I_{k_1}}, \text{ donde } C \text{ es una constante} \quad (1.8)$$

Ahora bien, antes de continuar, se define a x_{ij} , con $x_{ij} = 0$ o 1 , como la categoría elegida por el individuo i para la variable j , con $i \in \{1, I\}$ y $j \in \{1, J\}$, siendo I el número de individuos o registros y J el número de variables; y se considerará a K_j como el número de variables categóricas que tiene la variable categórica j . Así pues el número de individuos que llevan sólo una de las dos categorías es igual a:

$$I_{k_0 \neq k_1} = \sum_{i=1}^I (x_{ik_0} - x_{ik_1})^2$$

Entonces,

$$d_{k_0, k_1}^2 = C \frac{1}{I_{k_0} I_{k_1}} \sum_{i=1}^I (x_{ik_0} - x_{ik_1})^2, \quad (1.9)$$

$$= C \frac{1}{I_{k_0} I_{k_1}} \sum_{i=1}^I (x_{ik_0}^2 + x_{ik_1}^2 - 2x_{ik_0}x_{ik_1}) \quad (1.10)$$

$$d_{k_0, k_1}^2 = C \frac{1}{I_{k_0} I_{k_1}} \left(\sum_{i=1}^I x_{ik_0}^2 + \sum_{i=1}^I x_{ik_1}^2 - 2 \sum_{i=1}^I x_{ik_0}x_{ik_1} \right) \quad (1.11)$$

Recuerde que $x_{ik} = \{0, 1\}$, por lo tanto $x_{ik}^2 = x_{ik}$ y entonces $\sum_{i=1}^I x_{ik}^2 = \sum_{i=1}^I x_{ik} = I_k$, obteniendo

$$d_{k_0, k_1}^2 = C \left(\frac{1}{I_{k_1}} + \frac{1}{I_{k_0}} - 2 \frac{\sum_{i=1}^I x_{ik_0}x_{ik_1}}{I_{k_0} I_{k_1}} \right)$$

Note que $\frac{1}{I_k} = \frac{I_k}{I_k^2} = \frac{\sum_{i=1}^I x_{ik}^2}{I_k^2}$. Entonces se tiene finalmente que la distancia al cuadrado es,

$$d_{k_0, k_1}^2 = C \left(\frac{\sum_{i=1}^I x_{ik_1}^2}{I_{k_1}^2} + \frac{\sum_{i=1}^I x_{ik_0}^2}{I_{k_0}^2} - 2 \frac{\sum_{i=1}^I x_{ik_0} x_{ik_1}}{I_{k_0} I_{k_1}} \right) \quad (1.12)$$

$$= C \left(\sum_{i=1}^I \left(\frac{x_{ik_1}}{I_{k_1}} \right)^2 + \sum_{i=1}^I \left(\frac{x_{ik_0}}{I_{k_0}} \right)^2 - 2 \sum_{i=1}^I \left(\frac{x_{ik_0}}{I_{k_0}} \times \frac{x_{ik_1}}{I_{k_1}} \right) \right) \quad (1.13)$$

$$d_{k_0, k_1}^2 = C \sum_{i=1}^I \left(\frac{x_{ik_1}}{I_{k_1}} - \frac{x_{ik_0}}{I_{k_0}} \right)^2 \quad (1.14)$$

Ahora bien, como se mencionó antes, el MCA es una extensión del CA, en este último se define lo que es una tabla de contingencia, la cual consiste en una tabla que muestra los datos de las categorías de dos variables al mismo tiempo (Husson et al., 2011). Para una mejor comprensión de lo que es una tabla de contingencia véase la tabla 1.4

Tabla 1.4.

Ejemplo: Tabla de contingencia

		Sexo		
		Hombre	Mujer	Total
Área	Urbano	6837	6739	13576
	Rural	4411	4273	8684
Total		11248	11012	22260

En general, dentro de una tabla de contingencia con dos variables categóricas V_1 y V_2 para n individuos. Se llama x_{ij} al número de individuos que escogió las categorías i para V_1 (con I categorías) y j para V_2 (con J categorías). Luego, para obtener la suma de una categoría específica se fija la otra variable, esto se denota como:

$$x_{i\bullet} = \sum_{j=1}^J x_{ij}, \quad x_{\bullet j} = \sum_{i=1}^I x_{ij}, \quad n = x_{\bullet\bullet} = \sum_{i,j} x_{ij}.$$

En el análisis de correspondencias, se considera las tablas de probabilidad asociadas a tablas de contingencia, las cuales muestran la probabilidad de que un individuo lleve tanto la categoría i , como la j , el término general de esta probabilidad se describe como $f_{ij} = \frac{x_{ij}}{n}$, y también se define a los probabilidades marginales como:

$$f_{i\bullet} = \sum_{j=1}^J f_{ij}, \quad f_{\bullet j} = \sum_{i=1}^I f_{ij}, \quad f_{\bullet\bullet} = \sum_{i,j} f_{ij} = 1.$$

Regresando al análisis de correspondencias múltiples. Se define la distancia entre individuos, considerando $C = I/J$, de la siguiente manera

$$d_{i_0, i_1}^2 = \frac{I}{J} \sum_{k=1}^K \frac{1}{I_k} (x_{i_0 k} - x_{i_1 k})^2, \quad (1.15)$$

$$= \sum_{k=1}^K \frac{IJ}{I_k} \left(\frac{x_{i_0 k}}{J} - \frac{x_{i_1 k}}{J} \right)^2, \quad (1.16)$$

$$d_{i_0, i_1}^2 = \sum_{k=1}^K \frac{1}{I_k/(IJ)} \left(\frac{x_{i_0 k}/(IJ)}{1/I} - \frac{x_{i_1 k}/(IJ)}{1/I} \right)^2 \quad (1.17)$$

Al incluir las notaciones de la tabla de contingencia usada en CA, y aplicarla en una matriz de indicadores, se obtiene:

$$f_{ik} = x_{ik}/(IJ),$$

$$f_{\bullet k} = \sum_{j=1}^I x_{jk}/(IJ) = I_k/(IJ),$$

$$f_{i\bullet} = \sum_{k=1}^K x_{ik}/(IJ) = 1/I.$$

Una matriz de indicadores, es una matriz con los individuos en las filas y todas las categorías para cada variable en las columnas; en esta matriz x_{ik} toma el valor de 1 si el

individuo i cumple la categoría k , y 0 si no la cumple. Además, la matriz tiene dimensión $I \times K$ donde K satisface $K = \sum_{j=1}^J K_j$.

Así pues, se puede identificar la distancia χ^2 entre las filas i_0 e i_1 calculada a partir de la matriz de indicadores de la siguiente manera

$$d_{\chi^2}^2(\text{fila } i_0, \text{fila } i_1) = \sum_{k=1}^K \frac{1}{f_{\bullet k}} \left(\frac{f_{i_1 k}}{f_{i_1 \bullet}} - \frac{f_{i_0 k}}{f_{i_0 \bullet}} \right)^2 \quad (1.18)$$

Además, si se supone la constante $C = I$, la distancia al cuadrado entre dos categorías k_1 y k_0 se expresa como

$$d_{k_0, k_1}^2 = I \sum_{i=1}^I \left(\frac{x_{ik_1}}{I_{k_1}} - \frac{x_{ik_0}}{I_{k_0}} \right)^2$$

$$d_{k_0, k_1}^2 = \sum_{i=1}^I \frac{1}{1/I} \left(\frac{x_{ik_1}/(IJ)}{I_{k_1}/(IJ)} - \frac{x_{ik_0}/(IJ)}{I_{k_0}/(IJ)} \right)^2 \quad (1.19)$$

De esta manera se puede identificar la distancia χ^2 entre las columna k_1 y k_0 calculado a partir de la matriz de indicadores:

$$d_{\chi^2}^2(\text{fcolumna } k_0, \text{columna } k_1) = \sum_{i=1}^I \frac{1}{f_{i \bullet}} \left(\frac{f_{ik_1}}{f_{\bullet k_1}} - \frac{f_{ik_0}}{f_{\bullet k_0}} \right)^2 \quad (1.20)$$

Cabe recalcar lo enunciado por Husson et al. (2011), cuando dice que, en términos de cálculo, el MCA se basa en un análisis de correspondencia aplicado a una matriz de indicadores.

CAPÍTULO 2

2. METODOLOGÍA

2.1 Tratamiento de datos

En este proyecto se utilizaron bases de datos obtenidas de la página de la ENDI, pues están abiertas al público, por tal razón se mencionará a continuación un pequeño resumen de la metodología implementada por dicha institución. La ENDI pone a disposición del público las base de datos `BDD_ENDI_R1_f1_hogar.rds`, constando de 19187 filas y 119 columnas, donde recopila la información de los hogares donde viven niños menores de 5 años; `BDD_ENDI_R1_f1_personas.rds`, que consta de 90032 filas y 111 columnas, con información de todas las personas que viven en los hogares con niños menores de cinco años; `BDD_ENDI_R1_f1_mef.rds`, conteniendo 18955 filas y 275 columnas, con información sobre mujeres en edad fértil de diez a 49 años que tienen niños menores de 5 años; `BDD_ENDI_R1_f1_lactancia.rds`, con 11732 filas y 71 columnas e información sobre la lactancia de niños menores de tres años; y `BDD_ENDI_R1_f1_salud_ninez.rds`, que cuenta con 21333 filas y 412 columnas e información sobre la salud de los últimos dos niños nacidos vivos menores de cinco años que viven en los hogares registrados.

Todos estos datos fueron obtenidos a través del levantamiento de información hecho por personal del INEC mediante dos formularios. Las primeras dos bases de datos mencionadas

contienen la información distribuidas según las secciones del formulario 1 llamado "Formulario del Hogar". Mientras que las tres últimas bases de datos mencionadas contienen la información obtenida del formulario 2, según las secciones, llamado "Formulario Mujeres en Edad Fértil (MEF de 10 a 49 años), Salud de la Niñez y Lactancia Materna".

Cada una de las bases de datos tiene distintos tipos de variables, principalmente son de tipo identificadoras, categóricas o numéricas. Las variables presentes en las bases de datos que corresponden a preguntas de los formularios tienen la siguiente estructura: fm_sn_p , donde "f" y "s" son letras fijas que significan formulario y sección respectivamente, mientras que, m significa el número de formulario al que pertenece la variable, siendo estos 1 o 2, n varía según el número de sección de cada formulario al que pertenece la pregunta del formulario, y p es el número de pregunta que le corresponde en cada uno de los formularios.

Las distintas bases de datos contienen hasta siete tipos de variables identificadoras, según la vivienda, el hogar o la persona encuestada, si es mujer en edad fértil, si tiene hijos, entre otros. Estas variables si bien no brindan información relevante para el análisis, son de vital importancia para obtener información entre las bases de datos. Se tomó en consideración solo las siguientes variables identificadoras:

- id_hogar : El número de identificación asignado con base en la provincia, el cantón, la parroquia, el conglomerado, la vivienda y el hogar del entrevistado.
- id_per : El número de identificación derivado de $texttid_hogar$ asignado a las personas que viven en un hogar.
- id_mef : El número de identificación id_per asignado si la persona es una mujer en edad fértil.

Cabe mencionar el hecho de que en los formularios existen preguntas destinadas a distintas personas, como lo pueden ser el encargado de la casa, los niños menores a 5 años, las madres o las mujeres en edad fértil. Por tal razón fue necesario formular nuevas bases de datos agrupando variables de las bases de datos originales, es decir, sin alteración, en torno a los distintos grupos de personas. Por ejemplo, no brinda información relevante un registro de un hermano adolescente en el hogar, pues se consideró que no afecta directamente en la presencia o no de algún tipo de desnutrición infantil de los niños menores de cinco años.

Ahora bien, para la implementación de todos los procesos explicados a continuación se utilizó R como lenguaje de programación, empleando la interfaz Rstudio. En este proyecto se utilizaron los datos de `BDD_ENDI_R1_f1_personas.rds` y `BDD_ENDI_R1_f1_hogar.rds` para obtener la información de todos los niños menores de cinco años registrados. Se creó un dataframe llamado `ninosmenor5` en la cual se recopiló la información solo de los niños menores de 5 años, reduciendo así los registros de la base de datos `personas` de 90032 a 22260.

Una vez conformados los dataframes, se depuraron columnas en blanco tanto para los niños como para las mujeres en edad fértil, pues ciertamente los niños no tienen ningún registro de ingresos generados al hogar, y en el caso de las mujeres en edad fértil no tiene registro de las mediciones realizadas solo a los niños menores de 5 años.

Se realizó cruce de variables identificadoras mediante las variables `id_hogar`, `id_per`, `id_mef` y código mamá con el fin de obtener si las madres de los niños menores de cinco años vivían con ellos y así asignarles los valores según correspondía a los niños en la variable que guarda el nivel de educación de las madres. De igual manera, se realizó cruce de variables con la variable `id_hogar` entre las bases de datos de personas y hogar con el fin de asignarles a

los registros de los niños, la información de sus hogares, creando así un dataframe con nombre `DatosNiñosF`.

Para este estudio solo se tomarán en cuenta estas bases de datos, pues por la cantidad de registros y la presencia de valores N/A en las demás bases de datos, la población objetivo se reduciría considerablemente, además de pasar a ser; últimos dos niños nacidos vivos menores de cinco años por madres que viven con sus hijos.

2.1.1 Análisis de las variables

Posteriormente, se añadieron nuevas variables categóricas en el dataframe `DatosNiñosF`, con ayuda de documentos a disposición de la ENDI llamados Fichas técnicas, los cuales brindan información así como códigos de R para determinar niveles de anemia, nivel de cloro así como presencia o no de la bacteria *escherichia coli* en el agua y si tiene o no desnutrición crónica, global o aguda con base a los estándares de la OMS, esto último con ayuda del paquete `anthro` el cuál realiza los cálculos de puntuaciones z de antropometría infantil.

Los puntajes z antropometría infantil son valores que representan medidas que señalan la distancia y la dirección en la que se encuentra la medición de un individuo con respecto al valor promedio de las poblaciones de referencia. En el contexto de los estándares de crecimiento de la OMS, dicha población de referencia está constituida por niños del mismo género y rango de edad. Los puntajes Z que terminan fuera del rango normal indican problemas nutricionales. Cuanto más lejos se encuentre del rango normal, más grave llega a ser el problema nutricional (Cashin and Oot, 2018).

Así pues, se eliminaron del dataframes las variables utilizadas para determinar lo anterior mencionado, así como las variables identificadores pues estas no brindan información relevante,

y se añadieron las siguientes variables:

- `dcronica`: 1 si sufre desnutrición crónica, caso contrario 0.
- `daguda`: 1 si sufre desnutrición aguda, caso contrario 0.
- `dglobal`: 1 si sufre desnutrición global, caso contrario 0.
- `cloro_v/f`: 1 si un vaso/fuente del hogar presenta cloro, caso contrario 0.
- `ecoli_v/f`: 1 si un vaso/fuente del hogar presenta la bacteria escherichia coli, caso contrario 0.
- `prev_ane`: Determina el nivel de anemia según las categorías: No tiene anemia, Anemia leve, Anemia moderada, Anemia grave.

Se llamó `D1` al dataframe formado por 22260 filas y 59 columnas que contiene las variables mencionadas, además de variables sobre aspectos socioeconómicos, salud alimentaria, y condiciones de vida de los niños menores de 5 años. A continuación, se visualizó el dataframe para descubrir las posibles razones de los valores `N/A` y así completarlos, ya que esto es necesario para la implementación de los modelos. Así pues, se descubrió que los valores `N/A` de la variable `f1_s1_14`, correspondiente al tipo de establecimiento de enseñanza al que asiste el niño, son equivalentes a que el niño no está actualmente en alguna institución educativa. De igual manera, en la variable `nivins_mef`, nivel de educación de la madre del niño, los valores `N/A` son equivalentes a que la madre no vive en el hogar. En las demás variables donde se encontró valores `N/A` solo se completó con letras que significan registro sin datos o datos insuficientes, pues no se halló relación directa con alguna otra variable.

Previo a la implementación de los modelos, se crearon tres dataframes a partir de D1 llamados `dfDCsinna`, `dfDAsinna` y `dfDGsinna`, en los cuales se eliminaron los registros N/A de las columnas `dcronica`, `daguda` y `dglobal`, respectivamente. Adicionalmente en los dataframes `dfDAsinna` y `dfDGsinna` se eliminaron las variables que registran los otros dos tipos de desnutrición, pues la relación (en caso de existir) ya estaría reflejada en los resultados del dataframe `dfDCsinna`, ya que el modelo Decisión Tree es un método supervisado, es decir, se necesita una variable de respuesta.

2.2 Implementación de los modelos

A partir de lo mencionado anteriormente, se implementó, a cada uno de los dataframes, un modelo Decision Tree tomando como variable de respuesta, la presencia o no de desnutrición crónica, aguda y global, según correspondía. Con la finalidad de tomar sus medidas de acuracidad y kappa, para tener como referencia, al momento de implementar los respectivos Random Forest. Es decir no se enfocó en el árbol resultante, pues por su gran cantidad de variables, así como hojas, la interpretación gráfica de estos no brinda información clara.

Se dividió de forma aleatoria cada base de datos en dos conjuntos, un conjunto de entrenamiento y un conjunto de prueba, equivaliendo al 70% y 30% de los registros (filas) toda la base de datos. El conjunto de entrenamiento es empleado para la "creación" del modelo, mientras que el conjunto de prueba, es utilizado para probar el modelo de predicción creado a partir del conjunto de entrenamiento.

Obteniendo para la desnutrición crónica lo reflejado en la Figura 2.1, así como un valor **Accuracy** de 0.8414 y un **kappa score** igual a 0.2702. Luego, para la desnutrición aguda se obtuvo lo que muestra la Figura 2.2, con un **Accuracy** igual a 0.9927 y un valor **kappa** de 0. Finalmente,

se aplicó un árbol de decisión para el dataframe `dfDGsinna`, es decir, tomando como variable de respuesta la desnutrición global, y se obtuvo el árbol mostrado en la Figura 2.3, obteniendo un valor **Accuracy** de 0.9599 y un y un valor **kappa** igual a -0.0012

Figura 2.1.
Árbol de decisión Desnutrición Crónica

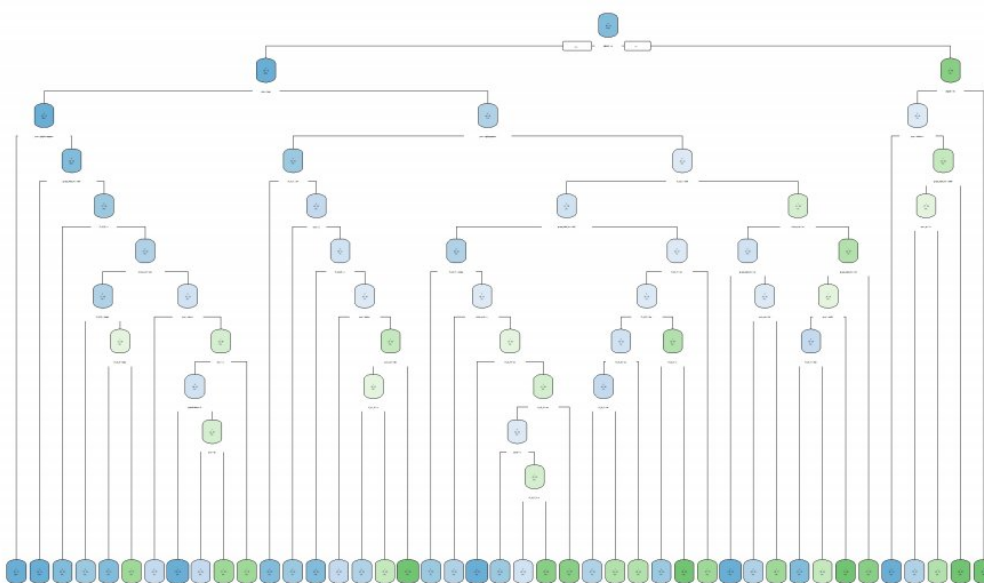
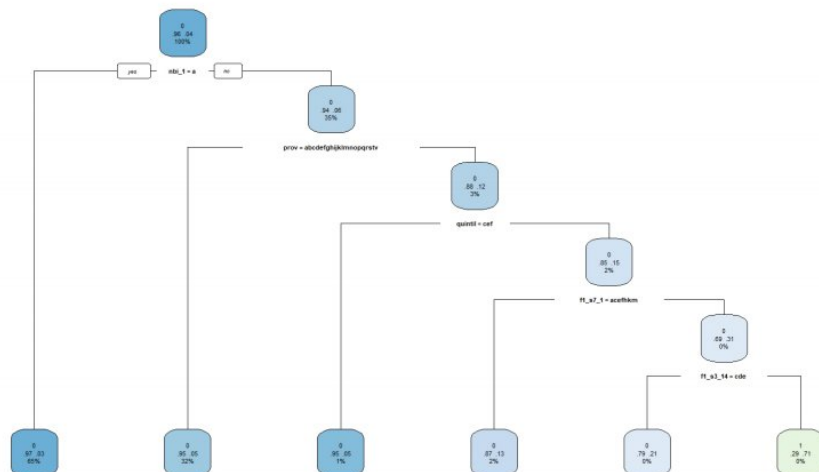


Figura 2.2.
Árbol de decisión Desnutrición Aguda



Posterior a esto, se implementó el modelo "Random Forest" a cada una de las bases de

Figura 2.3.
Árbol de decisión *Desnutrición Global*



datos, utilizando los mismos conjuntos de entrenamiento y prueba. Con el objetivo de determinar cuáles son las variables que tienen mayor influencia al momento de determinar si un niño tiene o no desnutrición infantil, según sea el tipo de desnutrición. Esta información se obtiene de dos formas, la primera dependiendo de que variables mejoran la exactitud del modelo (*Accuracy*) y la segunda dependiendo de que variables agrupan en mejor manera los datos (*Gini*). Es decir, este modelo brinda las variables más importantes al momento de determinar si un niño tiene desnutrición, obteniendo factores que afectan en la prevalencia de la desnutrición infantil.

Ahora bien, teniendo las variables que tienen un mayor peso a la hora de predecir, según el modelo, la presencia de desnutrición en los niños, se buscó determinar si existen relaciones entre estas variables. Para eso se aplicó la técnica MCA a los dos grupos de variables más importantes, según como mejorar el "accuracy" y el "gini", para cada tipo de desnutrición: crónica, aguda y global.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

El presente capítulo se divide en dos secciones, en las cuales se muestra los resultados de la aplicación del modelo Random Forest y la técnica MCA a cada una de las bases de datos. Cada sección está dividida en subsecciones según el tipo de desnutrición infantil. En la primera sección se detalla la precisión de cada Random Forest aplicado, así como las variables con mayor peso en la predicción. En la segunda sección se analizan los resultados de la implementación de los respectivos MCA, así como su reducción dimensional para con cada tipo de desnutrición

3.1 Resultados modelo Random Forest

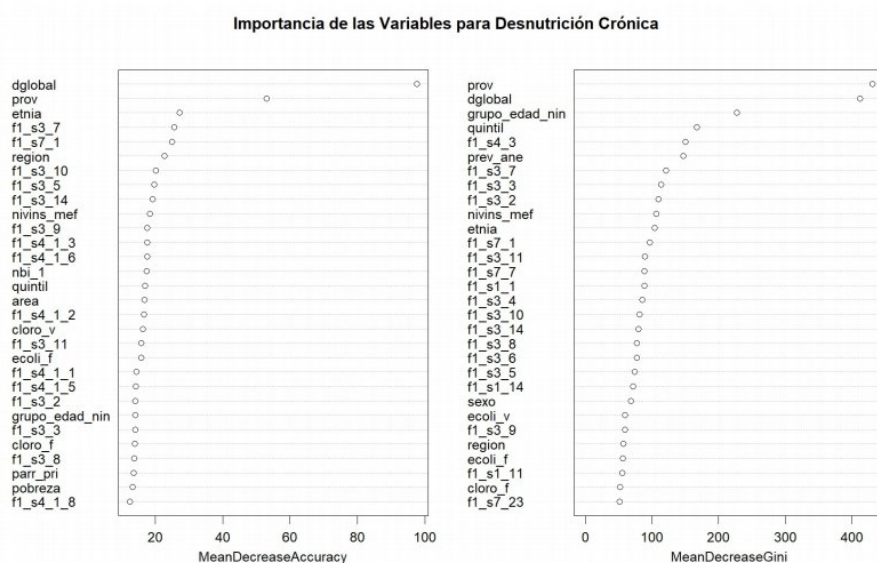
En esta subsección, se enunciará el principal resultado que fue de provecho para el estudio, como lo es la obtención de las variables que más afectan en la decisión de asignar a un niño menor de cinco años en el modelo, también llamadas las variables más importantes del modelo. De igual manera, se menciona los valores de accuracy y kappa en cada caso,

Cabe mencionar que los gráficos de esta subsección reflejan lo importante que es una variable para el algoritmo; es decir, mientras más arriba esté en el gráfico, quiere decir que si se llega a eliminar dicha variable, el modelo sería menos preciso o menos homogéneo en cada subdivisión, según corresponda.

3.1.1 Desnutrición Crónica

La implementación del modelo Random Forest a la base de datos *dfDCsinna*, siendo la variable de respuesta *dcronica*, la cual asigna 1 si el niño tiene desnutrición crónica, y 0 si no tiene, dio como resultado una **exactitud** del 84.45% y un valor de **kappa** igual a 0.2838, siendo este en su respectiva escala "fair". De igual manera, se obtiene que las variables con mayor importancia a la hora de determinar la presencia de desnutrición crónica son las que se pueden visualizar en la Figura 3.1, dependiendo tanto de la exactitud de predicción que brinda la variable (MeanDecreaseAccuracy), como la capacidad de agrupamiento de los datos que aporta dicha variable (MeanDecreaseGini).

Figura 3.1.
Importancia de variables para determinar DC según Accuracy y Gini



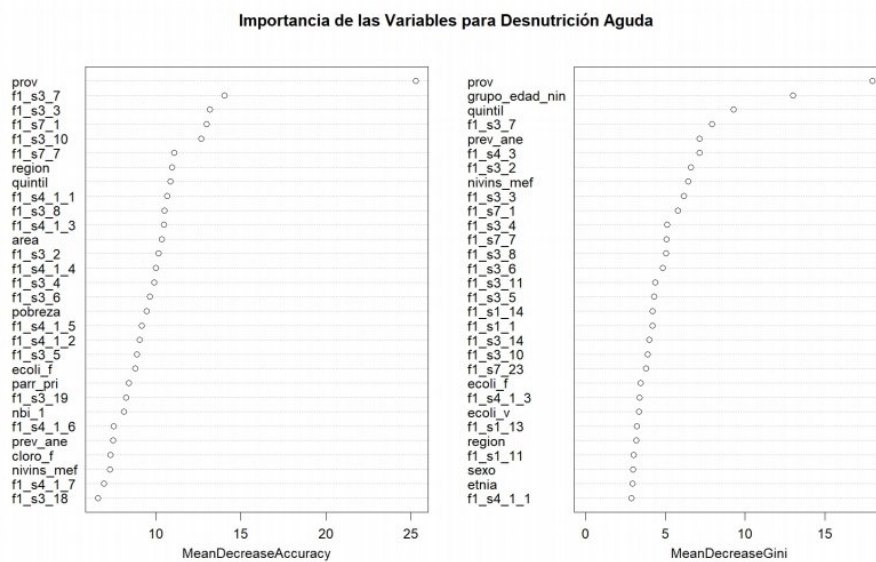
Así pues, se obtuvo una mejora en la exactitud y en el índice kappa de la predicción del modelo con respecto a un árbol de decisión. También se obtuvo que algunas de las variables que afectan a la presencia de desnutrición crónica, son la desnutrición global, la provincia y región a la

que pertenece, etnia del niño, el grupo de edad, el material del suelo de la casa, de donde proviene le agua que beben, entre otras.

3.1.2 Desnutrición Aguda

La implementación del modelo Random Forest a la base de datos dfDAsinna, siendo la variable de respuesta *daguda*, la cual asigna 1 si el niño tiene desnutrición aguda, y 0 si no tiene, dio como resultado una **exactitud** del 99.29% y un valor de **kappa** igual a 0.0414, siendo este en su respectiva escala "slight". También se obtuvo que las variables con mayor importancia a la hora de determinar la presencia de desnutrición aguda son las que se pueden visualizar en la Figura 3.2, dependiendo tanto de la exactitud de predicción que brinda la variable (MeanDecreaseAccuracy), como de la capacidad de agrupamiento de los datos que aporta dicha variable (MeanDecreaseGini).

Figura 3.2.
Importancia de variables para determinar DA según Accuracy y Gini



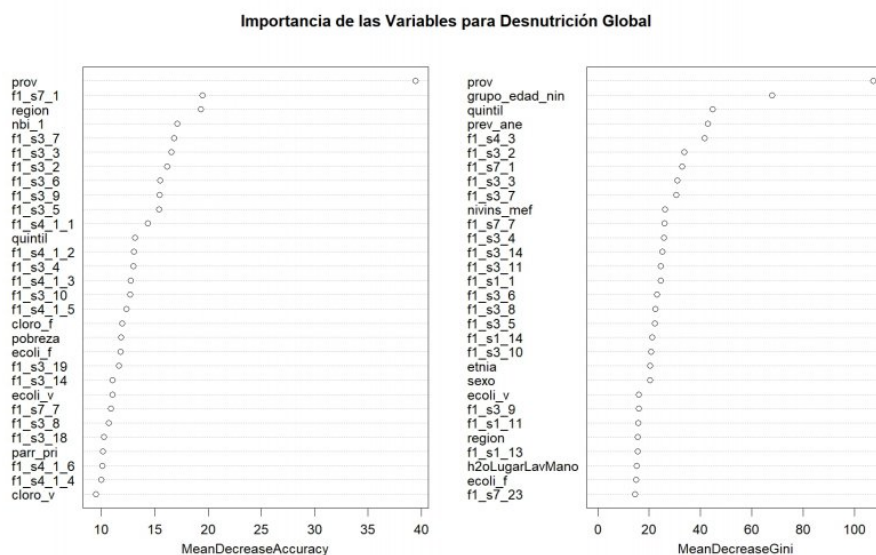
Así pues, se obtuvo una mejora casi completamente nula en la exactitud, mientras que en el índice kappa sí se notó alguna mejora de la predicción del modelo, pues con respecto al Decision Tree, esta era nula. Además se obtuvo que algunas de las variables que afectan a la presencia de desnutrición aguda en los niños menores de cinco años, son la provincia a la que pertenece, si el niño sufre de anemia y en qué nivel, el grupo de edad, el material del suelo y del techo de la casa, de dónde proviene el agua que beben, el quintil de pobreza al que pertenece su familia, entre otras.

3.1.3 Desnutrición Global

La implementación del modelo Random Forest a la base de datos `dfDGsinna`, siendo la variable de respuesta `dglobal`, la cual asigna 1 si el niño tiene desnutrición global, y 0 si no tiene, dio como resultado una **exactitud** del 96% y un valor de **kappa** igual a 0.0201, siendo este en su respectiva escala "slight". De igual manera, se obtuvo que las variables con mayor importancia a la hora de determinar la presencia de desnutrición global son las que se pueden visualizar en la Figura 3.3, dependiendo tanto de la exactitud de predicción que brinda la variable (`MeanDecreaseAccuracy`), como de la capacidad de agrupamiento de los datos que aporta dicha variable (`MeanDecreaseGini`).

Entonces, casi no se obtuvo mejora en la exactitud, mientras que en el índice kappa sí se notó alguna mejora de la predicción del modelo, pues con respecto al Decision Tree, esta era de nivel "poor". Además, se obtuvo que algunas de las variables que afectan a la presencia de desnutrición global en los niños menores de cinco años, son la provincia y región a la que pertenece, si el niño sufre de anemia y en qué nivel, el grupo de edad, el material del suelo, de donde proviene el agua que beben, el quintil de pobreza al que pertenece su familia, la percepción de la familia

Figura 3.3.
Importancia de variables para determinar DG según Accuracy y Gini



sobre la desnutrición infantil, entre otras.

3.2 Resultados MCA

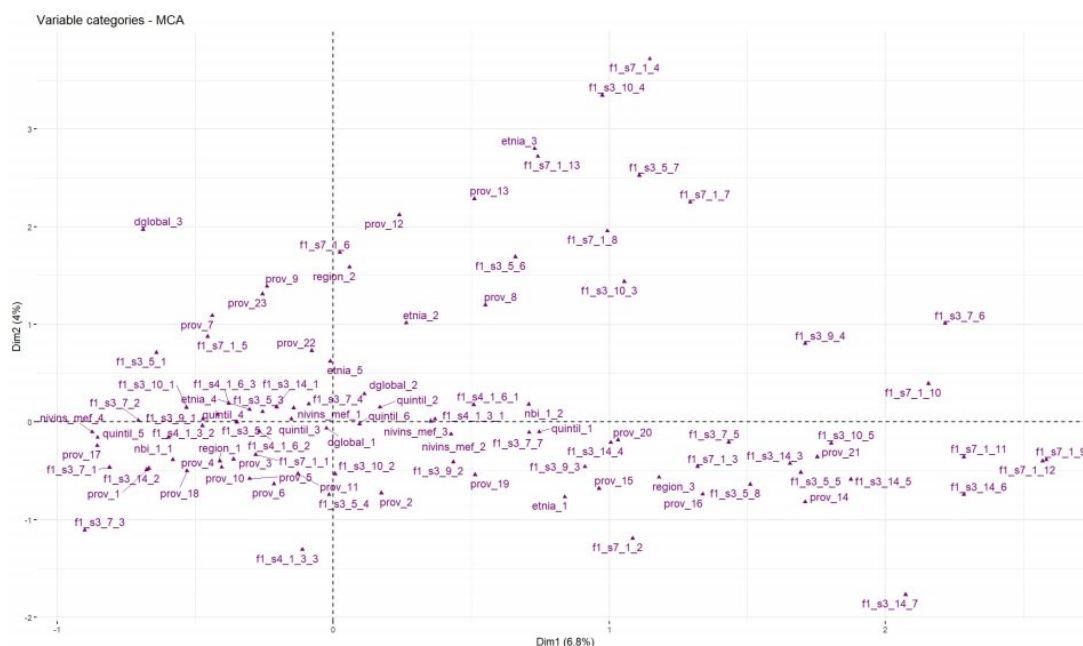
Ahora bien, se tomó un número de variables equivalente a más del 50% de importancia tanto para accuracy como para gini y se las aplicó solo a los registros de los niños quienes sufren el tipo de desnutrición correspondiente sea crónica, aguda y global; con el objetivo de determinar posibles relaciones entre las variables que más peso tuvieron en la predicción.

3.2.1 Desnutrición Crónica

Primero se muestra los resultados de las variables más importantes según la exactitud que le brindan al modelo Random Forest.

Como se observa en la Figura 3.4, las categorías de cada variable empleada están repartidas entre dos ejes, por tal razón se va a ver la contribución que hacen las categorías a

Figura 3.4.
Gráfico MCA con variables importantes para Accuracy en DCI



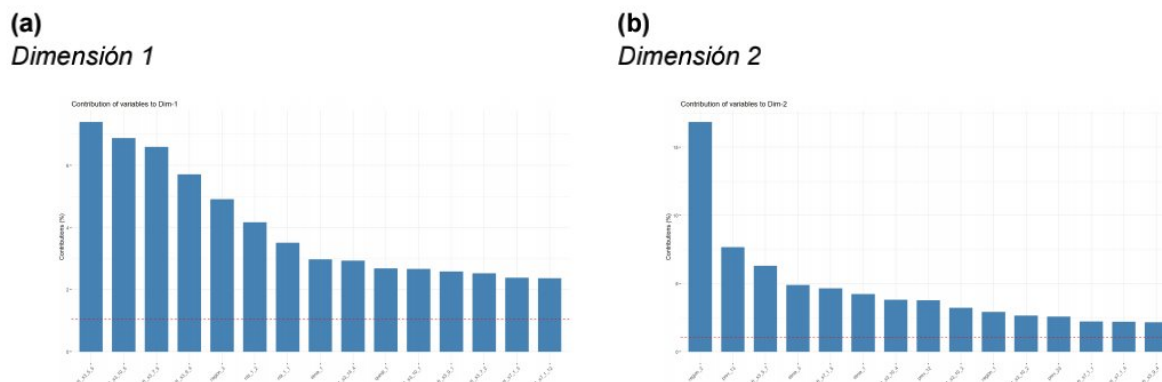
cada eje para poder determinar como se realizó a cabo la reducción dimensional.

Como indica la Figura 3.5, se obtuvo que la dimensión 1 está conformada por un conjunto de características más dispersas, siendo las cinco con mayor influencia en la dimensión: casas con paredes de madera, piso de tabla sin tratar, agua suministrada por fuentes naturales sin uso de tubería y vivir en la región Amazonía. Mientras que la dimensión 2 está más fuertemente relacionada con el hecho de vivir en la región Costa, vivir en una casa con pared de caña no revestida, o ser de la etnia montubia o indígena.

Aunque, cabe recalcar que la mayoría de las categorías tienen una incidencia en las dimensiones, solo que en mucho menor medida, pues el indicador de esto es la línea roja encontrada en los gráficos que indica el promedio de contribución entre todas las categorías.

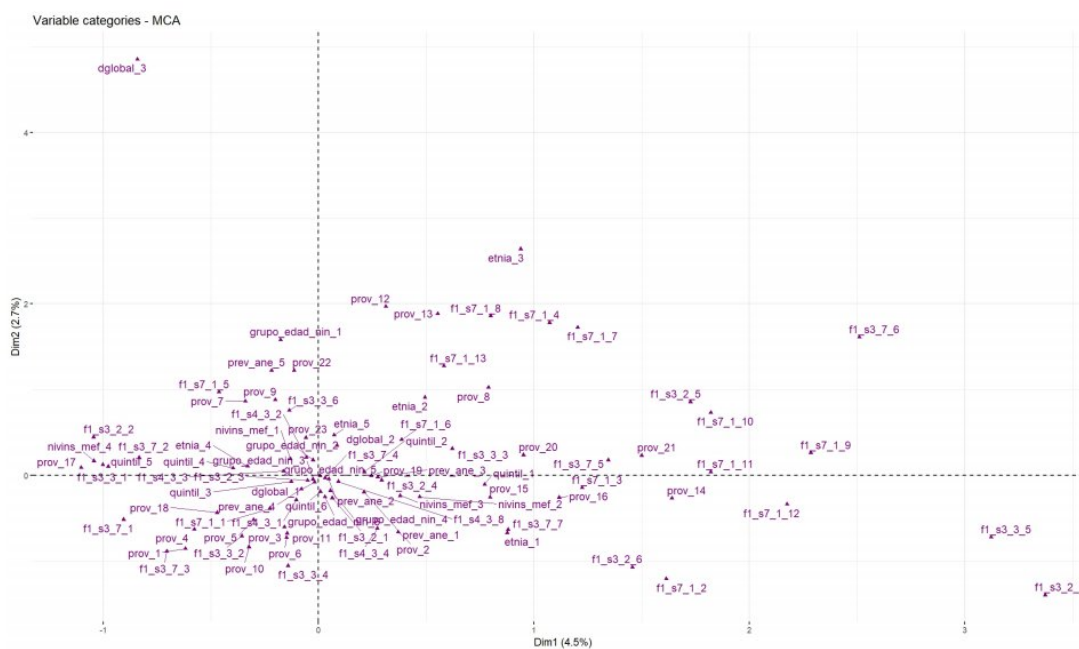
Procediendo de igual manera, pero con las variables más importantes según el gini se

Figura 3.5.
Contribución de las categorías de las variables para cada eje/dimensión



obtuvo lo que muestra la Figura 3.6. Así mismo se visualizó la contribución que hacen las categorías a cada eje para poder determinar cómo se llevó a cabo la reducción dimensional.

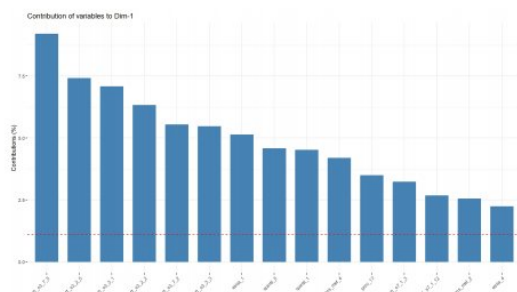
Figura 3.6.
Gráfico MCA con variables importantes para Gini en DCI



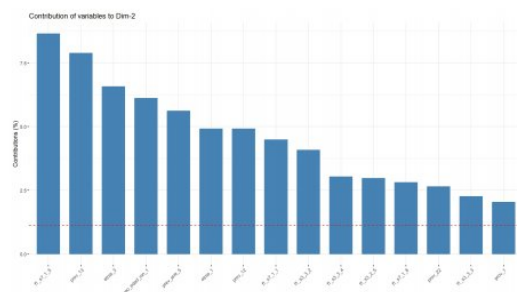
Como indica la Figura 3.7, se obtuvo que la dimensión 1 está conformado principalmente por los niños que cumplen ciertas características las cuales tiene una gran diversidad, siendo las

Figura 3.7.
Contribución de las categorías de las variables para cada eje/dimensión

(a)
Dimensión 1



(b)
Dimensión 2



cinco con mayor influencia en la dimensión: casas o tipo rancho o departamento, con suelo o de madera o de cerámica, con techos o de hormigón o de zinc, es decir un estilo de vida o campestre o urbano. Mientras que la dimensión 2 está más fuertemente relacionada con el hecho de beber agua de bidón, o pertenecer a un grupo de edad correspondiente a recién nacidos.

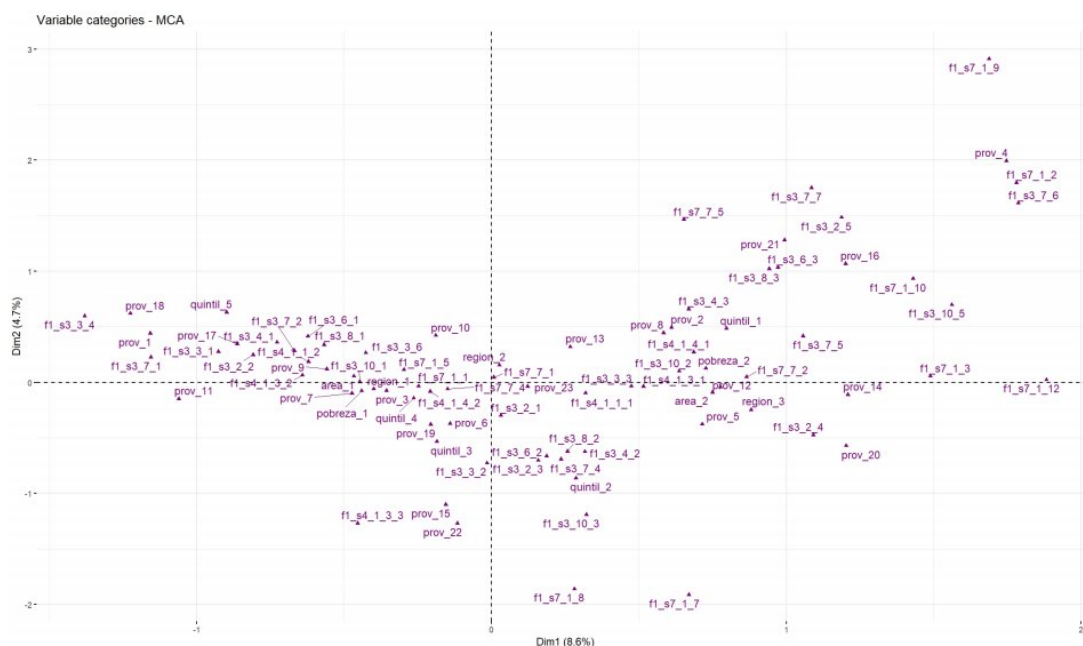
3.2.2 Desnutrición Aguda

Primero se muestra los resultados de las variables más importantes según la exactitud que le brindan al modelo Random Forest.

Como se ve en la Figura 3.8, las categorías de cada variable empleada están repartidas entre dos ejes, por tal razón se va a ver la contribución que hacen las categorías a cada eje para poder determinar como se llevó a cabo la reducción dimensional.

Como indica la Figura 3.9, se obtuvo que la dimensión 1 está conformado con un conjunto de diversas características, siendo las cinco con mayor influencia en la dimensión: obtener agua suministrada por fuentes naturales, pertenecer al sector rural, sufrir de pobreza por ingresos, vivir en casas con piso de tabla sin tratar. Mientras que la dimensión 2 está más fuertemente relacionada con el hecho de vivir en viviendas tipo rancho, con paredes en estado

Figura 3.8.
Gráfico MCA con variables importantes para Accuracy en DAI



regular o malo, vivir en una casa con piso de ladrillo en estado regular, o ser de la etnia montubia o indígena.

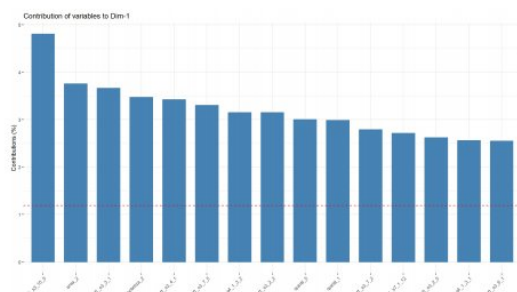
Aunque, cabe recalcar que la mayoría de las categorías tienen una incidencia en las dimensiones, solo que en mucho menor medida, pues el indicador de esto es la línea roja encontrada en los gráficos que indica el promedio de contribución entre todas las categorías.

Procediendo de igual manera, pero con las variables más importantes según el gini se obtuvo lo que muestra la Figura 3.10. Así mismo se visualizó la contribución que hacen las categorías a cada eje para poder determinar cómo se llevó a cabo la reducción dimensional.

Como indica la Figura 3.11, se obtuvo que la dimensión 1 está conformado principalmente por los niños menores de cinco que cumplen ciertas características, siendo las que tienen mayor influencia en la dimensión: casas con techos en buen estado de hormigón, así como un buen

Figura 3.9.
Contribución de las categorías de las variables para cada eje/dimensión

(a)
Dimensión 1



(b)
Dimensión 2

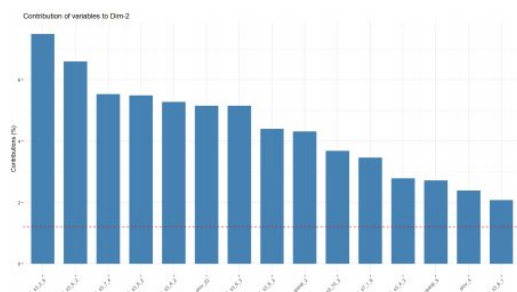
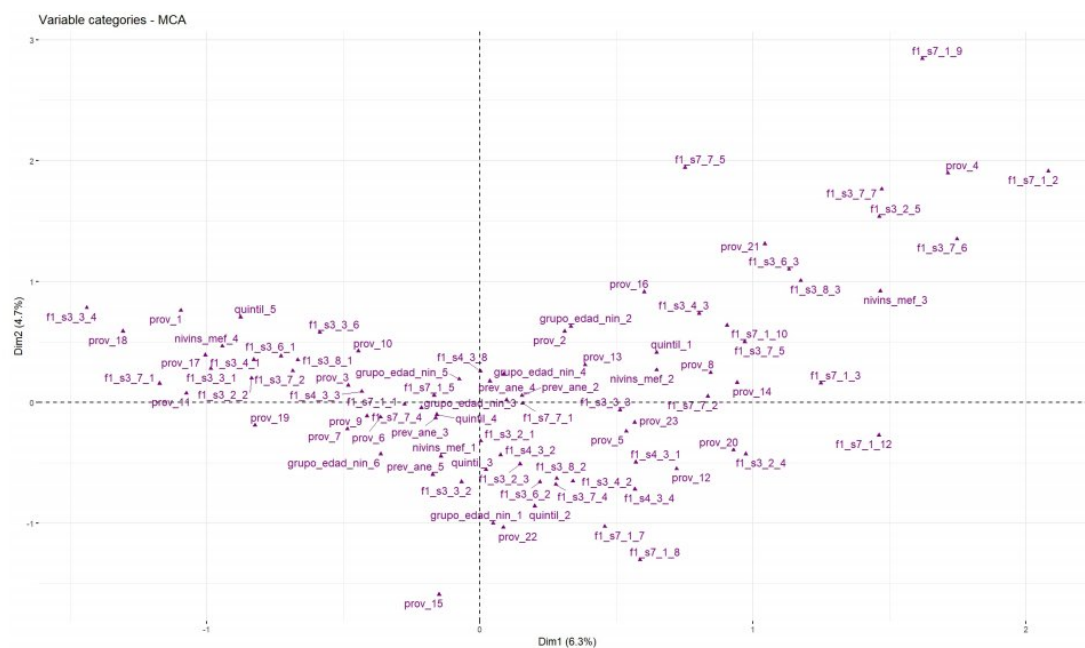


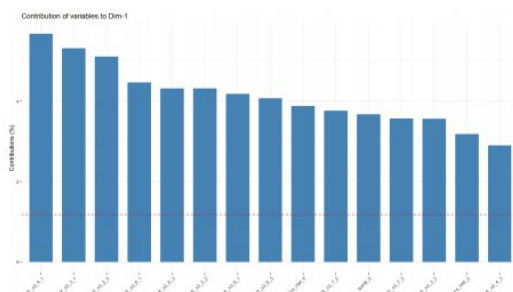
Figura 3.10.
Gráfico MCA con variables importantes para Gini en DAI



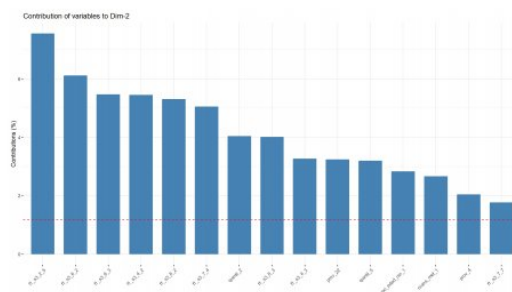
estado de las paredes exteriores y del piso, o el hecho de que las madres de los niños tenga una educación superior finalizada. Mientras que la dimensión 2 está más fuertemente relacionada con el hecho de tener paredes exteriores, piso y techo en estado regular o malo, pertenecer al segundo quintil de la población por ingresos.

Figura 3.11.
Contribución de las categorías de las variables para cada eje/dimensión

(a)
Dimensión 1



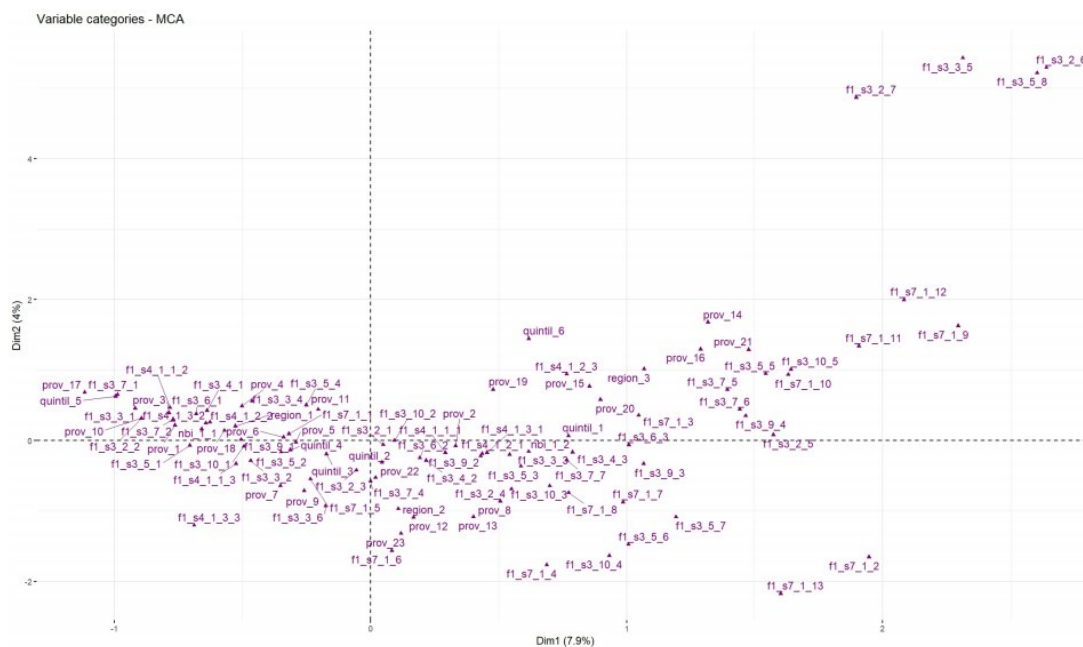
(b)
Dimensión 2



3.2.3 Desnutrición Global

Primero se muestra los resultados de las variables más importantes según la exactitud que le brindan al modelo Random Forest.

Figura 3.12.
Gráfico MCA con variables importantes para **Accuracy** en DGI



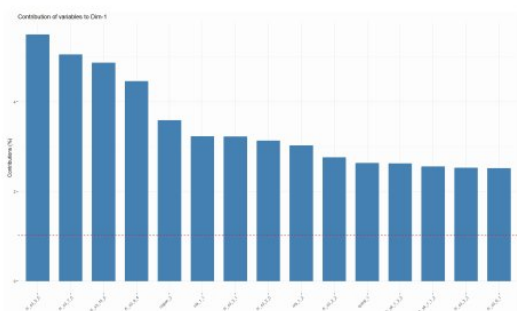
Como se observa en la Figura 3.12, las categorías de cada variable empleada están repartidas entre dos ejes, por tal razón se va a ver la contribución que hacen las categorías a cada eje para poder determinar cómo se llevó a cabo la reducción dimensional.

Figura 3.13.

Contribución de las categorías de las variables para cada eje/dimensión

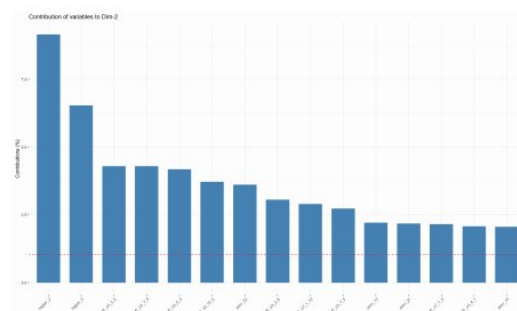
(a)

Dimensión 1



(b)

Dimensión 2

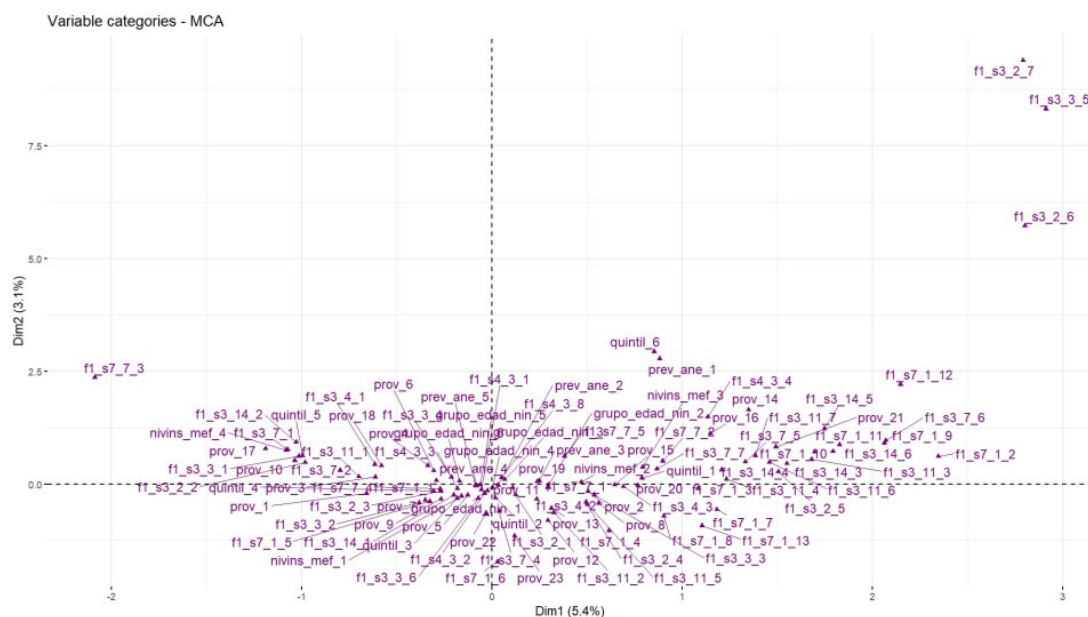


Como indica la Figura 3.13, se obtuvo que la dimensión 1 está conformado con un conjunto de diversas características, siendo las cuatro con mayor influencia en la dimensión: obtener agua suministrada por fuentes naturales, sin usar tuberías para acceder al agua, vivir en la región Amazónica o en casas con piso de tabla sin tratar y con paredes de madera. Mientras que la dimensión 2 está más fuertemente relacionada con el hecho de vivir en las regiones más tropicales del país .

Aunque, cabe recalcar que la mayoría de las categorías tienen una incidencia en las dimensiones, solo que en mucho menor medida, pues el indicador de esto es la línea roja encontrada en los gráficos que indica el promedio de contribución entre todas las categorías.

Procediendo de igual manera, pero con las variables más importantes según el gini se obtuvo lo que muestra la Figura 3.14. Así mismo se visualizó la contribución que hacen las categorías a cada eje para poder determinar como se realizó a cabo la reducción dimensional.

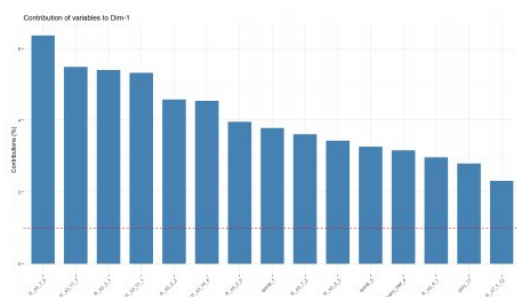
Figura 3.14.
Gráfico MCA con variables importantes para Gini en DGI



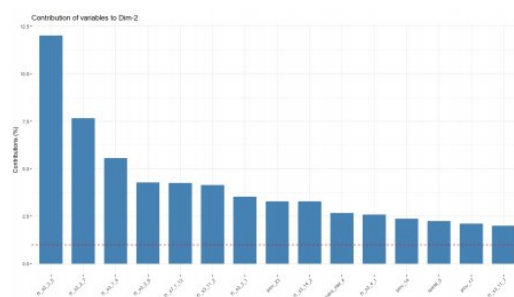
Como indica la Figura 3.15, se obtuvo que la dimensión 1 está conformado por una gran variedad de características que cumplen los niños menores de cinco años, siendo las que tienen mayor influencia en la dimensión: casas que o no tienen servicio higiénico o tienen inodoro conectado a red pública de alcantarillado, así como el hecho de quemar la basura del hogar. Mientras que la dimensión 2 está más fuertemente relacionada con el hecho de tener techos de plama, paja u hoja, vivir en viviendas tipo choza o covacha, beber agua de un río o acequia.

Figura 3.15.
Contribución de las categorías de las variables para cada eje/dimensión

(a)
Dimensión 1



(b)
Dimensión 2



CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

En este proyecto se investigaron ciertos factores que afectan a la prevalencia de la desnutrición infantil en los niños menores de cinco años en Ecuador. Se investigó la presencia de relaciones entre dichos factores para obtener información revelante que permita contribuir al cumplimiento del objetivo de desarrollo sostenible número dos. Se evaluó cómo el modelo Random Forest predice la presencia de desnutrición infantil en los niños menores de 5 años, comparando la precisión de este modelo con la precisión del modelo Decision Tree. Por último, se analizaron los resultados con el objetivo de establecer las relaciones existentes entre las variables que más afectan a la presencia de desnutrición infantil. A continuación, se presentan las conclusiones y recomendaciones más relevantes del proyecto.

Conclusiones

- Los factores que más contribuyen a la persistencia de cualquier tipo de desnutrición infantil en niños menores de 5 años son la prevalencia de anemia y sus niveles, el nivel de educación de las madres de los niños, así como la percepción que tienen ellas sobre la desnutrición infantil. También existen relaciones entre las provincias donde habitan y al grupo de edad al que pertenecen.
- La presencia de desnutrición crónica infantil se ve afectada principalmente por la presencia

de desnutrición global, la etnia a la que pertenece el niño menor de cinco años, el tipo de suelo predominante en sus hogares. Así mismo, dentro de los niños menores de cinco años que sí presentan desnutrición crónica se encontró que los hogares de los niños que beben agua de fuentes naturales de agua, están relacionados con los hogares que arrojan su basura a ríos, acequias, canales o quebradas. De igual manera se encontró una relación, dentro de los que cuentan con desnutrición infantil, entre las personas que viven en Morona Santiago y Orellana, con el hecho de que arrojan su basura a un terreno baldío o la queman, y consiguen el agua para sus hogares de cuerpos de agua.

- La presencia de desnutrición aguda y desnutrición global está más fuertemente afectada por las condiciones de obtención del agua que se usa en los hogares de los niños así como el agua que los niños menores de 5 años beben. De igual manera, en comparación con la desnutrición crónica, estas dos tienen mayor relación con el nivel de ingresos del hogar.

Recomendaciones

- Repetir este estudio con mejores medios computacionales, para poder realizar validación cruzada en los resultados, a fin de mejorar la precisión de las predicciones del modelo Random Forest de modo que se identifiquen más variables que puedan explicar de manera más efectiva la presencia de desnutrición infantil, además de poder visualizar de mejor manera una mayor cantidad de variables en los gráficos MCA para poder detectar mayor cantidad de relaciones entre las variables.
- En vista de que el grupo de edad al que pertenecen los niños es una variable que afecta de alguna manera la presencia de desnutrición infantil, se sugiere realizar un estudio enfocado

en los niños menores de tres años, pues la ENDI brinda una base de datos con más información sobre este grupo de edad, ofreciendo así la oportunidad de detectar nuevas variables que pueden afectar a la prevalencia de desnutrición infantil en niños menores de 3 años.

- Al usar modelos machine learning, existe la posibilidad constante de la aparición de mejores métodos, debido a esto se recomienda estudiar métodos más eficientes para establecer relaciones entre la presencia de desnutrición infantil y las demás variables que pone a disposición la ENDI, así como reducir la población para poder realizar un estudio más profundo sumando los datos de la base de datos que recopila la información de la salud de los dos últimos niños nacidos vivos menores de cinco años de cada hogar.

BIBLIOGRAFÍA

Aldás, J. and Uriel, E. (2017). *Análisis multivariante aplicado con R. 2ª ed.* Ediciones Paraninfo, S.A.

Banco Mundial (2021). El banco mundial y la nutrición. Obtenido de: <https://www.bancomundial.org/es/topic/nutrition/overview#1>.

Cashin, K. and Oot, L. (2018). *Guide to Anthropometry: A Practical Tool for Program Planners, Managers, and Implementers.* Washington, DC. Obtenido de : <https://www.fantaproject.org/sites/default/files/resources/FANTA-Anthropometry-Guide-May2018.pdf>.

Cuevas-Nasu, L., Gaona-Pineda, E. B., Rodríguez-Ramírez, S., Morales-Ruán, M. d. C., González-Castell, L. D., García-Feregrino, R., Gómez-Acosta, L. M., Ávila Arcos, M. A., Shamah-Levy, T., and Rivera-Dommarco, J. (2019). Desnutrición crónica en población infantil de localidades con menos de 100000 habitantes en México. *Salud Pública de México*, 61(6, nov-dic):833–840.

ENDI (2023). Boletín técnico nro. 01-2023-endi principales indicadores. Technical report, INEC.

ENDI (2023). Nota técnica para la comparación de la dci entre la endi 2022 – 2023 y la ensanut 2018 diseño muestral y calidad de los datos. Technical report, INEC.

FAO, FIDA, OPS, WFP, and UNICEF (2020). Panorama de la seguridad alimentaria y nutrición en América Latina y el Caribe. Santiago de Chile.

- Hair, J. (2010). *Multivariate Data Analysis*. Always learning. Prentice Hall.
- Husson, F., Lê, S., and Pagès, J. (2011). *Exploratory multivariate analysis by example using R*. Chapman and Hall, New York.
- INEC (2023a). Encuesta nacional sobre desnutrición infantil. Obtenido de: https://www.ecuadorencifras.gob.ec/encuesta_nacional_desnutricion_infantil/#camino.
- INEC (2023b). Principales resultados encuesta nacional sobre desnutrición infantil - endi.
- Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review* 39.
- Meneses, J. (2019). *Introducción al análisis multivariante*. FUOC, Barcelona.
- ODS Territorio Ecuador (2018). Logros y desafíos en la implementación de los ods en Ecuador. *Panorama Sostenible Anual No.1. Quito: ODS Territorio Ecuador*.
- Organización Mundial de la Salud (OMS) (s.f.). Health-topics malnutrition. Obtenido de: <https://www.who.int/health-topics/malnutrition>.
- Palma, A. (2018). Malnutrición en niños y niñas en América Latina y el Caribe. Obtenido de: <https://www.cepal.org/es/enfoques/malnutricion-ninos-ninas-america-latina-caribe>.
- Reeve, R. and Gottlieb, K. (2020). Sequentially determined measures of interobserver agreement (κ) in clinical trials may vary independent of changes in observer performance. *Therapeutic Innovation Regulatory Science*.
- Suthaharan, S. (2016). *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*. Springer US, Boston, MA.

Sánchez Hidalgo, M. d. R., Valdés Madrigal, I., González Fonseca, Z., Leyva Fonseca, D., and Figueroa Rodríguez, F. A. (2020). Factores socio-ambientales de riesgo de malnutrición por defecto. *Multimed*, 24:853 – 869.

UNICEF, World Health Organization, and International Bank for Reconstruction and Development/The World Bank (2023). *Levels and trends in child malnutrition: key findings of the 2023 edition of the joint child malnutrition estimates*.

Unicef Ecuador (2021). Desnutrición crónica infantil. uno de los mayores problemas de salud pública en Ecuador. Obtenido de: <https://www.unicef.org/ecuador/desnutrici%C3%B3n-cr%C3%B3nica-infantil>.

Vicepresidencia de la República del Ecuador (s.f.). Ecuador aprueba la endi, primera encuesta exclusiva del INEC para medir índices de desnutrición crónica infantil. Obtenido de: <https://www.vicepresidencia.gob.ec/ecuador-aprueba-la-endi-primer-encuesta-exclusiva-del-inec-para-medir-indices-de-desnutricion>.

WHO (2020). Q&A malnutrition. Obtenido de: <https://www.who.int/news-room/questions-and-answers/item/malnutrition>.

Zhou, Z.-H. (2021). *Decision Trees*, pages 79–102. Springer Singapore, Singapore.

APÉNDICES

APÉNDICE A

A.1 Significado de las variables

En esta sección se muestra una tabla que indica el significado de las variables mostradas en los resultados.

Tabla A.1.
Significado de las variables

Variables	Significado
prov	Provincia del ECuador
grupo_edad_nin	Grupo de edad al que pertenece el niño
quintil	Quintil de pobreza al que pertenece
prev_ane	Nivel de anemia que presenta el niño
region	Región geográfica del Ecuador
f1_s3_7	Material del piso de la vivienda
f1_s7_1	De dónde proviene el agua que beben
f1_s3_10	Quién suministra el agua que recibe la vivienda
f1_s3_5	Material de las paredes de la vivienda
f1_s3_14	Cómo elimina la basura de la vivienda
f1_s3_9	Medio por el que recibe el agua la vivienda
nivins_mef	Nivel de educación de la madre
f1_s4_3	Percepción sobre DCI en el hogar
f1_s3_3	Material del techo de la vivienda
f1_s3_2	Tipo de vivienda
f1_s3_11	Cómo es el servicio higiénico de la vivienda
f1_s4_1_3	¿El niño alguna vez comió poca variedad de alimentos por falta de recursos?
f1_s4_1_6	¿El hogar quedó alguna vez sin alimentos por falta de recursos?
nbi_1	Pobreza por necesidades básicas insatisfechas
f1_s1_1	Parentesco del niño con el representante del hogar