



Clasificación de la desnutrición e inseguridad alimentaria en población infantil de los hogares del departamento de Antioquia, Colombia: un enfoque utilizando aprendizaje automático.

Huberth Rolando Hincapié Arango

Yeisson Alejandro Gutiérrez Cano

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesores

Daniel Escobar Saltaren, Magíster en Ingeniería

Universidad de Antioquia

Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2024

Cita	(Hincapié Arango & Gutiérrez Cano, 2024)
Referencia Estilo APA 7 (2020)	Hincapié Arango, H. R. & Gutiérrez Cano, Y. A. (2024). <i>Clasificación de la desnutrición e inseguridad alimentaria en población infantil de los hogares del departamento de Antioquia, Colombia: un enfoque utilizando aprendizaje automático</i> . [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.



Especialización en Analítica y Ciencia de Datos, Cohorte VI.

Grupo de Investigación Intelligent Information Systems Lab – In2Lab.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Danny Alejandro Múnera Ramírez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

Este grandioso resultado lo dedico a Sara, Mateo y Silvana, mis hijos; quienes son el motor que impulsa las acciones importantes en mi vida; y a Stella y Alfonso, mis padres; por hacer germinar en mi la confianza, dedicación y enfoque indispensables para culminar con creces y llevar a buen término esta importante empresa.

Agradecimientos

Gracias a mi madre Emilse, a mi padre Fredy, a mis hermanos Farley, Kevin y Wayhonel, por ser mi sangre y el oxígeno de mi existir.

Contenido

I.	INTRODUCCIÓN.....	7
II.	MATERIALES Y MÉTODOS	9
	A. Fuente de datos.....	9
	B. Preprocesamiento de datos	9
	C. Codificación y selección de variables	10
	D. Imputación de datos	10
	E. Categorización de variables dependientes.....	11
	F. Entrenamiento de modelos	11
	G. Métricas de Rendimiento	11
	H. Validación Cruzada.....	12
III.	RESULTADOS Y DISCUSIÓN.....	13
	A. Variables seleccionadas.....	13
	B. Imputación de datos	14
	C. Construcción base de datos unificada	14
	D. Exploración de datos para O1 y O2	15
	E. Categorización de variables dependientes.....	15
	F. Construcción y métricas de los modelos	16
	G. Tuning de los modelos	16
IV.	CONCLUSIONES	21
V.	REFERENCIAS	22

Lista de Figuras

FIG. 1. PIVOTE DE LOS DATOS.....	10
FIG. 2. VALIDACIÓN CRUZADA: EVALUACIÓN DEL DESEMPEÑO DEL ESTIMADOR.....	12
FIG. 3. MATRIZ DE CORRELACIÓN PARA LAS VARIABLES PREDICTORAS	13
FIG. 4. PROCESO DE CONSOLIDACIÓN DE DATOS PARA CADA VARIABLE DESENLACE.....	15
FIG. 5. DISTRIBUCIÓN DE VARIABLES DESENLACE O1 Y O2	15
FIG. 6. FRECUENCIA DE LA INCIDENCIA DE DESNUTRICIÓN AGUDA EN MENORES DE 5 AÑOS SEGÚN NIVEL DE SEVERIDAD.....	15
FIG. 7. FRECUENCIA DE LA PREVALENCIA DE INSEGURIDAD ALIMENTARIA EN EL HOGAR SEGÚN NIVEL DE SEVERIDAD	16
FIG. 8. MATRIZ DE CONFUSIÓN EXTRA TREES CLASSIFIER O1	17
FIG. 9. ÁREA BAJO LA CURVA ROC O1	17
FIG. 10. MATRIZ DE CONFUSIÓN EXTRA TREES CLASSIFIER O2	18
FIG. 11. ÁREA BAJO LA CURVA ROC O2	18
FIG. 12. MATRIZ DE CONFUSIÓN RANDOM FOREST CLASSIFIER O1.....	19
FIG. 13. ÁREA BAJO LA CURVA ROC O1	19
FIG. 14. MATRIZ DE CONFUSIÓN RANDOM FOREST CLASSIFIER O2.....	20
FIG. 15. ÁREA BAJO LA CURVA ROC O2	20

Lista de Tablas

TABLA I. VARIABLES SELECCIONADAS	9
TABLA II. INDICADORES SELECCIONADOS	13
TABLA III. MÉTRICAS DE DESEMPEÑO DE LOS MODELOS PRIORIZADOS PARA CADA VARIABLE DESENLACE.....	16
TABLA IV. PARAMETROS PARA GRID SEARCH PARA EXTRA TREES CLASSIFIER O1	17
TABLA V. MÉTRICAS DE DESEMPEÑO EXTRA TREES CLASSIFIER O1	17
TABLA VI. PARAMETROS PARA GRID SEARCH PARA EXTRA TREES CLASSIFIER O2.....	18
TABLA VII. MÉTRICAS DE DESEMPEÑO EXTRA TREES CLASSIFIER O2	18
TABLA VIII. PARAMETROS PARA GRID SEARCH PARA RANDOM FOREST CLASSIFIER O1	19
TABLA IX. MÉTRICAS DE DESEMPEÑO RANDOM FOREST CLASSIFIER O1	19
TABLA X. PARAMETROS PARA GRID SEARCH PARA RANDOM FOREST CLASSIFIER O2	20
TABLA XI. MÉTRICAS DE DESEMPEÑO RANDOM FOREST CLASSIFIER O1	20

CLASIFICACIÓN DE LA DESNUTRICIÓN E INSEGURIDAD ALIMENTARIA EN POBLACIÓN INFANTIL DE LOS HOGARES DEL DEPARTAMENTO DE ANTIOQUIA, COLOMBIA: UN ENFOQUE UTILIZANDO APRENDIZAJE AUTOMÁTICO.

Resumen: La problemática de la desnutrición y la inseguridad alimentaria ha sido abordada desde diversos contextos: escolares, territoriales, sociodemográficos, y por medio de programas sociales. Así mismo, se han utilizado técnicas de análisis estadístico y predictivo para comprender los factores más relacionados con la problemática. En este artículo se presenta una estrategia para la construcción de bases de datos y se propone un modelo de machine learning capaz de clasificar la incidencia de desnutrición aguda en población menor de 5 años y la prevalencia de inseguridad alimentaria en los hogares del departamento de Antioquia. El modelo ha sido entrenado a partir de 67 indicadores recopilados por la Secretaría Seccional de Salud y Protección Social de Antioquia. Se trabajó con métricas comunes como: Accuracy, área bajo la curva ROC (AUC), recall, precisión, F1-Score, coeficiente kappa, Coeficiente de Correlación de Matthews (MCC) y matriz de confusión. Se compararon 16 modelos diferentes para cada indicador; entre los cuales el Random Forest Classifier presentó la mayor precisión. Luego, se realizó el tuning de este modelo para cada indicador, obteniendo un ROC de 84% para el indicador de inseguridad alimentaria y 79% para desnutrición infantil.

Palabras clave: *desnutrición, inseguridad alimentaria, aprendizaje automático, políticas públicas.*

I. INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) plantea que una buena nutrición influye positivamente en la salud de los lactantes, así como en embarazos y partos más seguros [1]. Según esta organización, los niños desnutridos son más vulnerables ante la enfermedad y la muerte, 149 millones de niños menores de 5 años en todo el mundo presentaban retraso del crecimiento en 2022 y casi la mitad de las defunciones en esta población tienen que ver con la desnutrición [2]. Por otra parte, la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) evidencia un aumento en la insuficiencia de consumo habitual de alimentos y la inseguridad alimentaria en el mundo desde 2014, proponiendo la ejecución de acciones urgentes [3].

Actualmente, alrededor del 38.8% de la población mundial enfrenta hambre crónica e inseguridad alimentaria moderada o grave [4]. En Colombia entre 2005 y 2022 se presentaron en promedio 372 muertes anuales por desnutrición en población menor de 5 años [5]. En otros países de

América Latina, la desnutrición infantil cobra la vida de más de 6,800 niños al día; y está asociado con problemas en el desarrollo físico y cognitivo, afectando el aprendizaje, el pensamiento y el habla [6]; además de reducir la capacidad intelectual, influye sobre patrones de comportamiento durante la adultez [7], afectando la productividad económica de los individuos y las sociedades futuras [8].

En Antioquia, las desigualdades sociales y económicas están relacionadas con alta prevalencia de inseguridad alimentaria en los hogares, habitar áreas rurales y vivir en condiciones socioeconómicas deterioradas [9]. Solo en la capital del departamento de Antioquia, para la década de 2011 a 2021, se dio una leve disminución del 3% en la desnutrición infantil [10], una cifra muy baja si se busca generar menos vulnerabilidad en la población.

La problemática de la desnutrición y la inseguridad alimentaria ha sido abordada desde diversos contextos: escolares [11], territoriales [12], sociodemográficos [8], [13], y por medio de programas sociales [14]. Así mismo, se han utilizado técnicas de análisis estadístico y predictivo para comprender los factores más relacionados con la problemática. Hay reportes de estudios que han implementado modelo univariante [7], modelos de regresión logística y lineal [8], [10], [15], árboles de decisión [6], [16], análisis de correspondencia Múltiple [17] y bosques aleatorios

[18]. Destacando que los modelos con mejor capacidad de predicción alcanzaron 95.00%, 90.03% y 69.76%, para regresión logística, Random Forest y AdaBoost, respectivamente.

El entrenamiento de este tipo de modelos puede, por lo tanto, ser utilizado por los diferentes “tomadores de decisión” u otros *stakeholders* que podrían servirse de un sistema de predicción y clasificación del riesgo de desnutrición alimentado por IA [19], [20]. Este tipo de modelos serviría también para realizar simulaciones de escenarios en los cuales se atacan los factores con mayor incidencia en el riesgo de desnutrición, posiblemente permitiendo una mejor planeación de las políticas públicas o planes de intervención social.

Para poder determinar los factores predominantes en la desnutrición infantil y la inseguridad alimentaria a nivel regional, en este proyecto se propone un modelo de aprendizaje automático para la clasificación de la incidencia de desnutrición en población infantil y la prevalencia de la inseguridad alimentaria en los hogares del territorio Antioqueño. El modelo ha sido entrenado a partir de 98 indicadores recopilados por la Secretaría Seccional de Salud y Protección Social de Antioquia.

Para ello en la sección de materiales y métodos se presentan la fuente de datos, su preparación y segmentación en los indicadores de desnutrición e inseguridad alimentaria en la

población de interés y se presentan las métricas de desempeño utilizadas para seleccionar los modelos más prometedores. Posteriormente se realiza una comparación entre 5 técnicas de IA diferentes, a partir de las cuales se seleccionan las más precisas para realizar un proceso de optimización de hiperparámetros. Finalmente, se concluye sobre la combinación de técnicas que mejor logren modelar la incidencia de desnutrición infantil, la prevalencia de inseguridad alimentaria y sus posibles factores causantes.

II. MATERIALES Y MÉTODOS

Los contenidos del desarrollo del proceso metodológico, datos y obtención de resultados pueden encontrarse en el repositorio de GitHub con acceso público [21].

A. Fuente de datos

La principal fuente de datos se originó en el Servicio de Información y Estadística de la Secretaría Seccional de Salud y Protección Social de Antioquia [22], [23], el cual pone a disposición con acceso libre y seguro datos con resultados de los indicadores de desnutrición y seguridad alimentaria, además de otras características poblacionales [24] y estadísticas vitales [25].

El conjunto de datos contiene 135,386 instancias y 98 indicadores enmarcados en los Objetivos de Desarrollo Sostenibles (ODS), con desagregación para 125 municipios del departamento de Antioquia para el periodo

comprendido entre los años 2007 y 2022. Adicionalmente, esta información asume diferentes escenarios de uso y rendición de cuentas públicas, que están gobernados por leyes dentro del margen de política pública, al igual que otras leyes de orden presupuestal [22], [24], [25].

B. Preprocesamiento de datos

La fuente de datos presenta 18 variables para la identificación de cada indicador poblacional. Las variables utilizadas para proceder con el análisis propuesto son las que identifican el municipio, el año y la medición para cada indicador, y se listan en la siguiente (ver **TABLA I**). Las otras 13 variables segmentan la información por subregión u otros datos geográficos, además de información sobre los objetivos de desarrollo sostenibles a los cuales corresponden los indicadores en cuestión.

TABLA I. VARIABLES SELECCIONADAS

Variable	Descripción
Código entidad	Código municipio nomenclatura estandarizada DANE
Entidad	Nombre del municipio o entidad territorial
Año	Año al que corresponde la medición del indicador
Indicador	Nombre del indicador
Valor	Medición del indicador

Inicialmente la estructura de los datos identificaba en cada registro la medición y las características de cada indicador, con el objetivo de obtener la estructura de datos idónea para el entrenamiento de los modelos planteados, se realizó una transformación conocida como pivoteo [26], la cual reorganiza el dataframe basada en los atributos del nombre del indicador, generando al final un conjunto de datos con 1,753 instancias y 101 columnas, en la cual cada columna cuenta con las variables predictoras y las dependientes (ver Fig. 1).

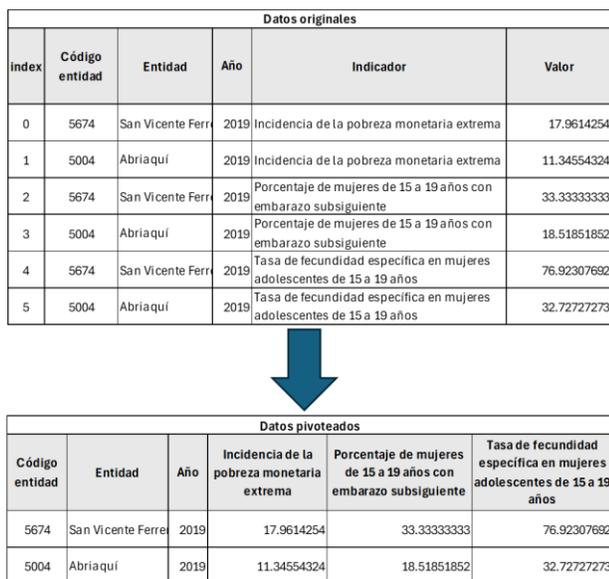


Fig. 1. Pivoteo de los datos

C. Codificación y selección de variables

Se realiza la codificación mapeando desde V0 hasta V95 (para los 96 indicadores) y las variables de salida O1 y O2. O1 corresponde a la incidencia de desnutrición aguda en población menor de 5 años y O2 a la prevalencia de

inseguridad alimentaria en el hogar (moderada o severa).

Basados en el Coeficiente de correlación de Person [27], [28], se hizo la selección de las variables con una correlación moderada, débil o nula (índice < 0.6), se excluyeron las variables para las cuales se obtuvo una correlación fuerte [29].

D. Imputación de datos

Con la intención de optimizar el reemplazo de los valores faltantes en las mediciones de los indicadores poblacionales como variables predictoras, se eligió el método de imputación K-nearest Neighbors [30]. Haciendo uso de la clase KNNImputer de la biblioteca scikit-learn disponible en Python y utilizando un parámetro n_neighbors de 15, con el fin de captar mejor la estructura subyacente de los datos, se tomaron en cuenta los 15 vecinos más cercanos para la imputación de cada valor faltante, y un parámetro de ponderación de los vecinos en función de la distancia, donde los valores de los vecinos más cercanos tienen un mayor impacto en el valor imputado que los valores de los vecinos más lejanos [31], [32].

Con el objetivo de depurar los registros con datos nulos en las variables dependientes (O1 y/O2) con respecto a las variables predictoras, se excluyeron el 28.7% de las instancias.

E. Categorización de variables dependientes

La medición de los indicadores desenlace O1 y O2, al ser inicialmente cuantitativa, no permite llevar a cabo el proceso de modelamiento para su clasificación, por ello se procedió a categorizarlas a partir de los cuartiles Q1, Q2 y Q3 en los niveles de Leve, Moderado y Alto, segmentados en el 33% y 66% de los datos respectivamente, esto como valor final para la magnitud de cada indicador en clases ya balanceadas [33].

F. Entrenamiento de modelos

Se realizó una separación en conjunto de entrenamiento y validación para la fase exploratoria inicial. Luego, para experimentar y explorar de manera eficiente la construcción de los modelos de machine learning y poder comparar ágilmente entre diferentes técnicas y resultados, se usó la biblioteca de código abierto de Python PyCaret 3.0, permitiendo obtener métricas de desempeño para 16 modelos diferentes [34]. Posteriormente se priorizaron, en función de la precisión (Accuracy), utilizando la curva ROC como métrica de desempate, los 5 mejores modelos para proceder a priorizar 2 para su despliegue y optimización de parámetros. Para la optimización de parámetros se utilizó un grid-search a partir del cual se variaron los parámetros más determinantes para las diferentes técnicas seleccionadas [35].

G. Métricas de Rendimiento

La precisión (Accuracy), es la métrica preferida por los investigadores tanto en problemas dicotómicos como multiclase, ya que es fácil de calcular y entender, lo que la convierte en una herramienta eficaz para evaluar el rendimiento general del algoritmo. Su forma de cálculo se basa en la proporción de clasificaciones predichas de manera correcta sobre el total de instancias [36].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Otra métrica empleada es la curva ROC, que es un gráfico bidimensional que representa la sensibilidad frente a cada clase. La métrica utilizada para la comparación en este gráfico es el AUC, que corresponde al área bajo la curva ROC y cuyos valores varían entre 0 y 1. Su forma de cálculo se basa en la probabilidad de clasificar correctamente una clase positiva al azar más que una negativa escogida al azar [36].

$$AUC = \frac{\text{recall} - (1 - \text{especificidad}) + 1}{2}$$

El recall, o sensibilidad, es una métrica que indica la proporción de casos positivos correctamente clasificados. En un modelo ideal, el recall es igual a 1 para cada clase. Desde una perspectiva analítica, un investigador busca incrementar el recall sin comprometer la exactitud. Su forma de cálculo se basa en la proporción de casos positivos bien clasificados [36].

$$Recall = \frac{TP}{TP + FN}$$

La precisión mide la proporción de verdaderos positivos sobre el total de predicciones positivas, una precisión alta se relaciona con una tasa baja de falsos positivos, e indica cuántas de las predicciones positivas del modelo son realmente correctas [37].

$$Precision = \frac{TP}{TP + FP}$$

El puntaje F1-Score es la media armónica de la precisión y el recall, y resulta especialmente útil cuando se trabaja con una distribución desigual de clases, este proporciona un balance entre ambas métricas [37].

$$F1 - Score = 2 \frac{Precision \cdot Recall}{Precision + Recall}$$

El coeficiente kappa es una métrica que evalúa la concordancia entre dos observadores. Este coeficiente ajusta la tasa de accuracy observada considerando el acuerdo que podría ocurrir por azar con la tasa de accuracy que se esperaría [36].

$$K = \frac{Po - Pe}{1 - Pe}$$

El Coeficiente de Correlación de Matthews (MCC) evalúa la eficacia de un clasificador binario, independientemente del tamaño de la muestra. Un MCC cercano a 1 indica una excelente clasificación, 0 indica que la clasificación no es mejor que una predicción aleatoria, y -1 significa que la clasificación es completamente opuesta a los resultados reales [38].

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TN + FP)(TP + FN)(TN + FN)(TP + FP)}}$$

H. Validación Cruzada

Se utilizó la validación mediante Kfold con el objetivo de verificar la capacidad de generalización del modelo. Se dividieron los datos en 10 subconjuntos (o folds) de manera aleatoria, garantizando la reproducibilidad de los resultados mediante una semilla aleatoria fija, y permitiendo, mediante dicha configuración, evaluar el rendimiento del modelo de manera más robusta y menos sesgada al utilizar diferentes combinaciones de datos de entrenamiento y prueba en cada iteración del proceso de validación cruzada [39], [40]. La figura 2 evidencia el funcionamiento de la validación cruzada con el uso de Kfold.

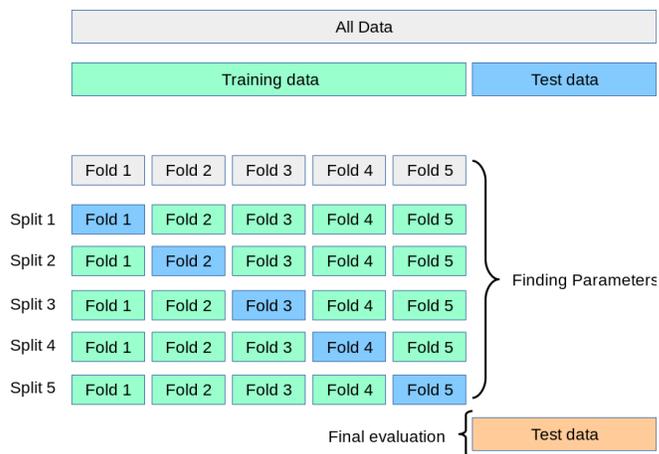


Fig. 2. Validación cruzada: evaluación del desempeño del estimador

Nota: fuente https://scikit-learn.org/stable/modules/cross_validation.html. User Guide, Model selection and evaluation.

Este procedimiento en la aplicación de Kfold, se presenta para todas las técnicas entrenadas en la sección de resultados.

III. RESULTADOS Y DISCUSIÓN

A. Variables seleccionadas

Se excluyeron 29 variables las cuales obtuvieron una correlación fuerte, y que para criterio obtuvieron un coeficiente de correlación de Pearson mayor a 0.6. Posterior a la depuración de dichas variables, se obtuvo una matriz de correlación que reflejó en un mapa de calor la moderada, débil o nula correlación entre las variables definitivas (ver **Fig. 3**), las cuales representan un total de 67 indicadores poblacionales como posibles predictores de la incidencia de desnutrición aguda en población menor de 5 años y/o la prevalencia de inseguridad alimentaria en el hogar (moderada o severa).

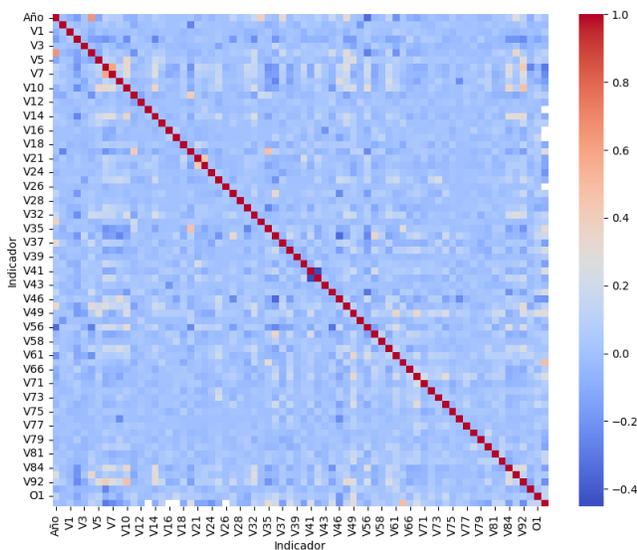


Fig. 3. Matriz de correlación para las variables predictoras

A continuación (ver **TABLA II**), se listan los indicadores poblacionales definidos para la construcción de los modelos de machine learning como variables predictoras.

TABLA II. INDICADORES SELECCIONADOS

Código	Nombre
V0	Brecha de ingreso mensual promedio entre hombres y mujeres
V1	Brecha de pobreza extrema urbano-rural
V2	Brecha entre cobertura neta urbano-rural en educación básica y media
V3	Capacidad instalada de generación de energía eléctrica
V4	Cobertura de establecimientos financieros
V5	Cobertura de vacunación con BCG en nacidos vivos
V6	Cobertura de viviendas con agua potable
V7	Cobertura de viviendas con agua potable (rural)
V9	Cobertura de viviendas con energía eléctrica
V10	Cobertura en educación superior
V11	Coefficiente de GINI
V12	Crecimiento anual del Producto Interno Bruto (PIB) real per-cápita
V13	Entidades territoriales con planes territoriales de cambio climático
V14	Exportaciones totales como porcentaje del Producto Interno Bruto (PIB)
V15	Hectáreas de pequeña y mediana propiedad rural, formalizadas
V16	Hogares urbanos con déficit cualitativo de vivienda
V17	Hogares urbanos con déficit cuantitativo de vivienda
V18	Incidencia de intoxicaciones por sustancias químicas
V19	Incidencia de la pobreza monetaria extrema
V21	Incidencia de violencia intrafamiliar (Atención x 2 fuentes)
V22	Incidencia de violencia psicológica (Atención x 2 fuentes) - SIVIGILA en solo mujeres
V24	Intensidad en el uso del agua per-cápita
V25	Intensidad energética
V26	Miles de hectáreas de áreas marinas protegidas
V27	Miles de hectáreas de áreas protegidas
V28	Mortalidad departamental causada por eventos recurrentes
V31	Mujeres que habitan en hogares donde hay teléfono celular
V32	Negocios verdes verificados
V33	Participación de mujeres en cargos de elección popular
V35	Porcentaje de jóvenes que no estudian y no tienen empleo
V36	Porcentaje de la superficie cubierta por bosque natural
V37	Porcentaje de menores de 1 año con tercera dosis de pentavalente

Código	Nombre
V38	Porcentaje de mujeres de 15 a 19 años con embarazo subsiguiente
V39	Porcentaje de mujeres en cargos directivos, según municipios de Antioquia
V40	Porcentaje de mujeres que han experimentado alguna violencia física por una persona diferente al esposo o compañero
V41	Porcentaje de mujeres que han sido forzadas físicamente por el esposo o compañero a tener relaciones o actos sexuales
V42	Porcentaje de mujeres que han sido forzadas físicamente por otra persona diferente a su esposo o compañero a tener relaciones o actos sexuales
V43	Porcentaje de municipios que cuentan con mecanismo intersectorial para el abordaje de las violencias de género
V45	Porcentaje de niños y niñas en primera infancia que cuentan con las atenciones priorizadas en el marco de la atención integral
V46	Porcentaje de partos atendidos por personal calificado
V47	Porcentaje de personas que requieren y son atendidas al menos una vez en los servicios en salud mental
V49	Porcentaje de población afiliada al sistema de seguridad social en salud
V55	Razón de mortalidad materna
V56	Sedes educativas conectadas a internet
V57	Superficie deforestada
V58	Superficie regenerada de bosque natural
V60	Tasa de cobertura bruta en educación media
V61	Tasa de cobertura bruta en transición
V63	Tasa de desempleo
V66	Tasa de homicidio de mujeres (ocurrencia del hecho)
V69	Tasa de mortalidad en menores de 5 años
V71	Tasa de mortalidad perinatal
V72	Tasa de mortalidad por VIH/SIDA
V73	Tasa de mortalidad por accidentes de tránsito terrestre
V74	Tasa de mortalidad por desnutrición en menores de 5 años
V75	Tasa de mortalidad por lesiones auto infligidas intencionalmente (suicidio)
V76	Tasa de mortalidad por malaria
V77	Tasa de mortalidad por tuberculosis
V78	Tasa de mortalidad por tumores malignos en menores de 18 años
V79	Tasa de mortalidad prematura de las enfermedades crónicas de las vías respiratorias inferiores
V80	Tasa de mortalidad prematura por cáncer de mama, cuello uterino, colon y recto, pulmón, próstata, estómago, linfoma no Hodgkin y Leucemia en adultos

Código	Nombre
V81	Tasa de mortalidad prematura por diabetes
V82	Tasa de mortalidad prematura por enfermedades cardiovasculares
V84	Tasa de solicitudes de patentes presentadas por vía nacional
V90	Vehículos eléctricos e híbridos matriculados
V92	Índice de bancarización
V95	Índice de riesgo de la calidad del agua (IRCA) Urbano

B. Imputación de datos

El proceso de imputación realizado sobre las 67 variables predictoras permitió obtener un conjunto de datos almacenado en nuevo dataframe, el cual conservo los índices y columnas del dataframe original, conteniendo la totalidad de datos para las variables predictoras posterior al proceso de imputación, igualmente, se evidencian datos faltantes en las variables O1 y O2, para lo cual se llevó a cabo un proceso de depuración de registros con datos nulos para cada indicador desenlace de forma independiente.

C. Construcción base de datos unificada

A partir del conjunto de datos obtenido posterior al proceso de pivoteo, selección e imputación de variables, el proceso de depuración de valores nulos se llevó a cabo para cada indicador desenlace de forma independiente, obteniendo dos conjuntos de datos, uno para la incidencia de desnutrición y otro para la prevalencia de inseguridad alimentaria, con 750 y 500 registros respectivamente (**Fig. 4**).



Fig. 4. Proceso de consolidación de datos para cada variable desenlace

D. Exploración de datos para O1 y O2

La depuración de los registros con datos faltantes en los indicadores de la incidencia de desnutrición aguda en población menor de 5 años y la prevalencia de inseguridad alimentaria en el hogar (moderada o severa), permitió contar finalmente con mediciones reales para O1 en los años de 2017 a 2022, y para O2 en los años de 2013, 2017, 2019 y 2021.

Estos indicadores presentan una distribución con valores de medición a partir de cero y con datos extremos a los cuales no se les realizó manejo como atípicos, ya que corresponden a mediciones del indicador dentro de un contexto dinámico poblacional, y reflejan el comportamiento del evento en los territorios sujetos de dicha medición (Fig. 5).

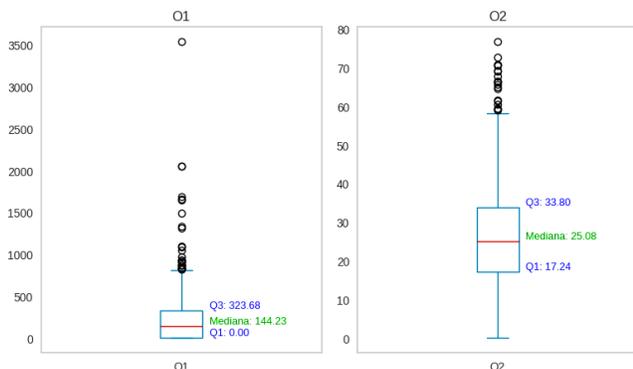


Fig. 5. Distribución de variables desenlace O1 y O2

E. Categorización de variables dependientes

La categorización de la incidencia de desnutrición aguda en población menor de 5 años permitió conocer el comportamiento de dicho indicador desde tres niveles de severidad, en un nivel leve con un rango que va desde 0 hasta 63.5 casos por 100 mil menores de 5 años, moderado hasta 240.5 casos por 100 mil menores de 5 años y alto hasta 3,533.6 casos por 100 mil menores de 5 años. En términos de frecuencia, el nivel alto presenta picos de mayor incidencia en la población de estudio (Fig. 6).

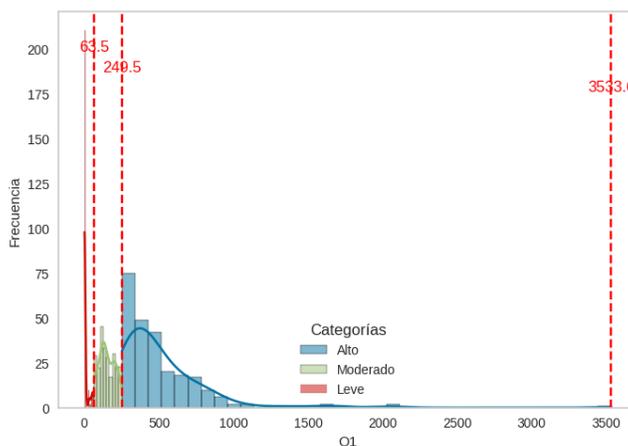


Fig. 6. Frecuencia de la incidencia de desnutrición aguda en menores de 5 años según nivel de severidad

La categorización de la prevalencia de inseguridad alimentaria en el hogar (moderada o severa) permitió conocer el comportamiento de dicho indicador desde tres niveles de severidad, en un nivel leve con un rango que va desde 0 hasta 19.7%, moderado hasta 30.7% y alto hasta 76.6%. En términos de frecuencia, el nivel alto presenta

picos de mayor prevalencia en la población de estudio (**Fig. 7**).

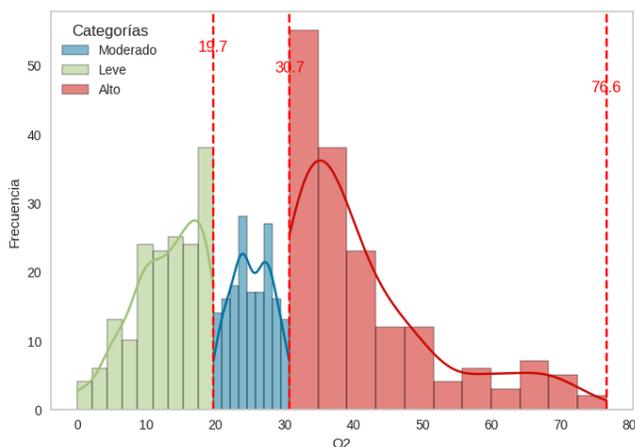


Fig. 7. Frecuencia de la prevalencia de inseguridad alimentaria en el hogar según nivel de severidad

F. Construcción y métricas de los modelos

Mediante el uso de PyCaret fue posible establecer y visualizar la configuración del entorno de experimentación para cada indicador desenlace. El entrenamiento se llevó a cabo con un total de 10 folds, 525 instancias para O1 y 350 instancias para O2, y el testeo con 225 y 150 instancias para O1 y O2 respectivamente. El resto de las métricas fueron usadas en su configuración por defecto.

G. Tuning de los modelos

Mediante PyCaret se realizó la priorización de 3 modelos con los mejores desempeños entre 16 comparados, obteniendo para cada indicador O1 y O2 las siguientes métricas (ver **TABLA III**).

TABLA III. MÉTRICAS DE DESEMPEÑO DE LOS MODELOS PRIORIZADOS PARA CADA VARIABLE DESENLACE

Desenlace	Modelo		Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Incidencia de desnutrición aguda en población menor de 5 años	et	Extra Trees Classifier	0.5602	0.7389	0.5602	0.561	0.556	0.3403	0.3429	0.248
	rf	Random Forest Classifier	0.5427	0.7281	0.5427	0.5461	0.5375	0.314	0.3181	0.355
	catboost	CatBoost Classifier	0.522	0.7241	0.522	0.5324	0.5213	0.2833	0.2865	35.875
	xgboost	Extreme Gradient Boosting	0.5143	0.701	0.5143	0.5203	0.5128	0.2714	0.274	1.282
	lightgbm	Light Gradient Boosting Machine	0.5087	0.7053	0.5087	0.5105	0.5066	0.263	0.2645	2.478
Prevalencia de inseguridad alimentaria en el hogar (moderada o severa)	rf	Random Forest Classifier	0.6486	0.8104	0.6486	0.6625	0.6473	0.4722	0.4775	0.558
	et	Extra Trees Classifier	0.64	0.8099	0.64	0.651	0.6403	0.4593	0.4628	0.227
	xgboost	Extreme Gradient Boosting	0.6314	0.8085	0.6314	0.6414	0.6311	0.4465	0.4502	0.913
	lightgbm	Light Gradient Boosting Machine	0.6286	0.8079	0.6286	0.6406	0.6288	0.4426	0.4464	0.654
	catboost	CatBoost Classifier	0.6143	0.8137	0.6143	0.6252	0.6125	0.4208	0.4259	30.901

1) *Extra Trees Classifier variable O1:*

Para la *incidencia de desnutrición aguda en menores de 5 años*, se llevó a cabo el entrenamiento del algoritmo de clasificación basado en árboles de decisión Extra Trees Classifier con los siguientes parámetros (ver **TABLA IV**). El entrenamiento se llevó a cabo con 10 pliegues para cada 2,160 combinaciones de parámetros, totalizando 21,600 entrenamientos.

TABLA IV. PARAMETROS PARA GRID SEARCH PARA EXTRA TREES CLASSIFIER O1

Parámetro	Valores	Mejores parámetros
n_estimators	[100, 200, 300]	200
max_depth	[5, 10, 15, 20, 25]	20
min_samples_split	[1, 2, 5, 10]	2
min_samples_leaf	[1, 2, 4]	1
max_features	[auto, sqrt, log2]	auto
bootstrap	[False, True]	False
criterion	[gini, entropy]	entropy
random_state	[135]	135

La matriz de confusión del modelo de clasificación mostró 47 verdaderos positivos para los niveles de incidencia de la desnutrición leve, moderado y alto, y un total de 28 falsos positivos (ver **Fig. 8**). La accuracy del modelo fue del 63%, con una precisión de 61%, 67% y 60% para leve, modera y alto respectivamente. La medida F1-Score, que equilibra precisión y exhaustividad, alcanzó un 0.65 para leve y moderado y 0.56 para alto, y el Recall 0.71 (leve), 0.64 (moderado) y 0.52 (alto) (**TABLA V**).

TABLA V. MÉTRICAS DE DESEMPEÑO EXTRA TREES CLASSIFIER O1

O1	Precisión	Recall	F1-Score
Alto	0.60	0.52	0.56
Leve	0.61	0.71	0.65
Moderado	0.67	0.64	0.65

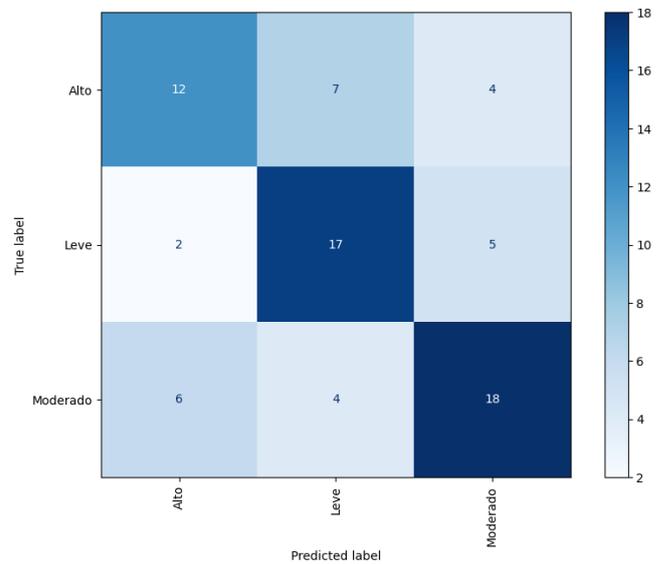


Fig. 8. Matriz de confusión Extra Trees Classifier O1

El AUC del modelo de clasificación es de 0.84 para una clasificación leve, 0.83 para moderado y 0.73 para un nivel alto (ver **Fig. 9**).

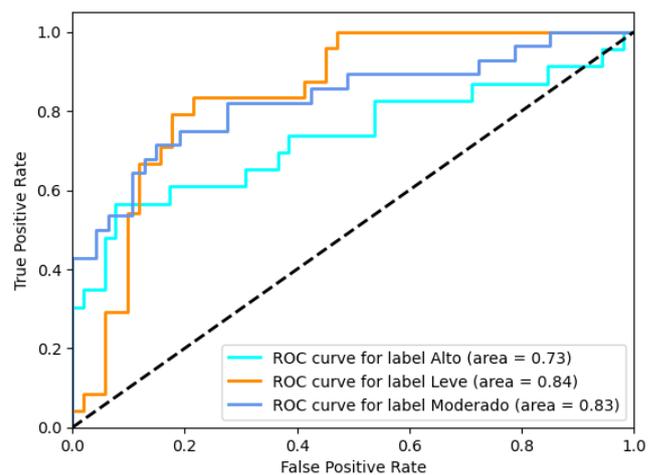


Fig. 9. Área bajo la curva ROC O1

2) *Extra Trees Classifier variable O2:*

Para la *prevalencia de inseguridad alimentaria en el hogar*, se llevó a cabo el entrenamiento del algoritmo de clasificación

basado en árboles de decisión Extra Trees Classifier con los siguientes parámetros (ver **TABLA VI**). El entrenamiento se ejecutó con 10 pliegues para cada 768 combinaciones de parámetros, para un total de 7,680 entrenamientos.

TABLA VI. PARAMETROS PARA GRID SEARCH PARA EXTRA TREES CLASSIFIER O2

Parámetro	Valores	Mejores parámetros
n_estimators	[300, 350, 400]	300
max_depth	[20, 25, 30, 40]	20
min_samples_split	[2, 3, 4, 5]	4
min_samples_leaf	[1, 2]	1
max_features	['sqrt', 'log2']	sqrt
bootstrap	[False, True]	True
criterion	['gini', 'entropy']	entropy
random_state	[135]	135

La matriz de confusión del modelo de clasificación mostró 30 verdaderos positivos para los niveles de prevalencia de inseguridad alimentaria leve, moderado y alto, y un total de 20 falsos positivos (ver **Fig. 10**). La accuracy del modelo fue del 60%, con una precisión de 73%, 47% y 62% para leve, moderada y alto respectivamente. La medida F1-Score, que equilibra precisión y exhaustividad, alcanzó un 0.65 para leve, 0.53 para moderado y 0.62 para alto, y el Recall 0.58 (leve), 0.60 (moderado) y 0.52 (alto) (**TABLA VII**).

TABLA VII. MÉTRICAS DE DESEMPEÑO EXTRA TREES CLASSIFIER O2

O2	Precisión	Recall	F1-Score
Alto	0.62	0.62	0.62
Leve	0.73	0.58	0.65
Moderado	0.47	0.60	0.53

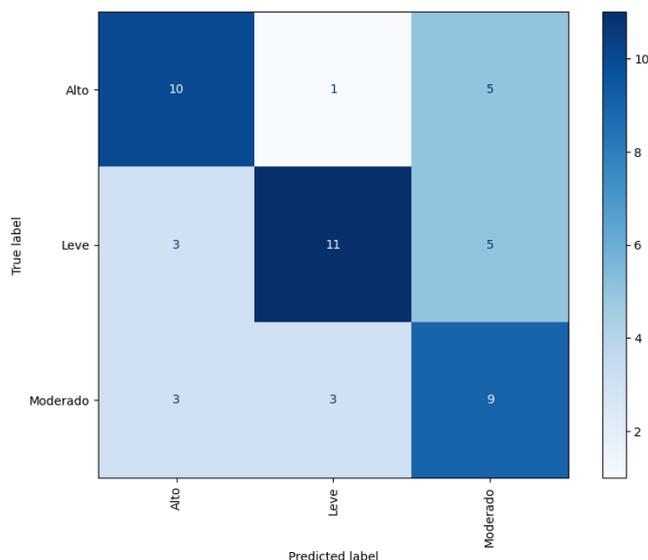


Fig. 10. Matriz de confusión Extra Trees Classifier O2

El AUC del modelo de clasificación es de 0.83 para una clasificación leve, 0.74 para moderado y 0.84 para un nivel alto (ver **Fig. 11**).

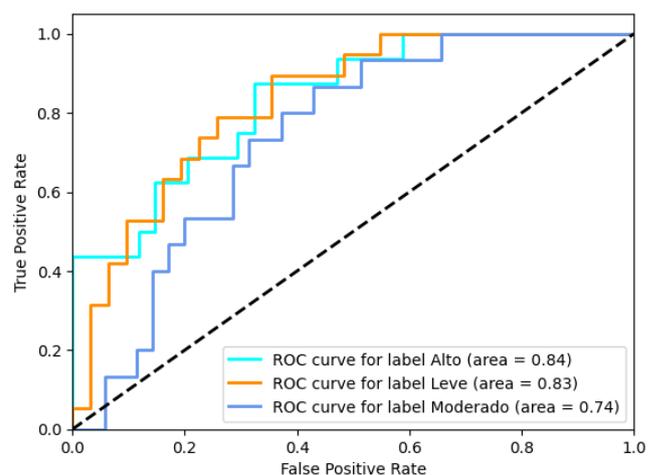


Fig. 11. Área bajo la curva ROC O2

3) Random Forest Classifier variable O1:

Para la incidencia de desnutrición aguda en menores de 5 años, se llevó a cabo el entrenamiento del algoritmo de clasificación basado en árboles de decisión Random Forest Classifier con los siguientes parámetros (ver **TABLA VIII**). Se entrenó con 10 pliegues para cada una de las 432 combinaciones de parámetros, para un total de 4,320 entrenamientos.

TABLA VIII. PARAMETROS PARA GRID SEARCH PARA RANDOM FOREST CLASSIFIER O1

Parámetro	Valores	Mejores parámetros
n_estimators	[100, 200]	1001
max_depth	[10, 20]	20
min_samples_split	[2, 5, 10]	5
min_samples_leaf	[1, 2, 4]	1
max_features	['auto', 'sqrt', 'log2']	log2
bootstrap	[False, True]	False
criterion	['gini', 'entropy']	entropy
random_state	[135]	35

La matriz de confusión del modelo de clasificación mostró 49 verdaderos positivos para los niveles de incidencia de la desnutrición leve, moderado y alto, y un total de 26 falsos positivos (ver Fig. 12). La accuracy del modelo fue del 65%, con una precisión de 68% para leve y 64% para moderado y alto. La medida F1-Score, que equilibra precisión y exhaustividad, alcanzó un 0.69 para leve, 0.64 para moderado y 0.62 para alto, y el Recall 0.71 (leve), 0.64 (moderado) y 0.61 (alto) (TABLA IX).

TABLA IX. MÉTRICAS DE DESEMPEÑO RANDOM FOREST CLASSIFIER O1

O1	Precisión	Recall	F1-Score
Alto	0.64	0.61	0.62
Leve	0.68	0.71	0.69
Moderado	0.64	0.64	0.64

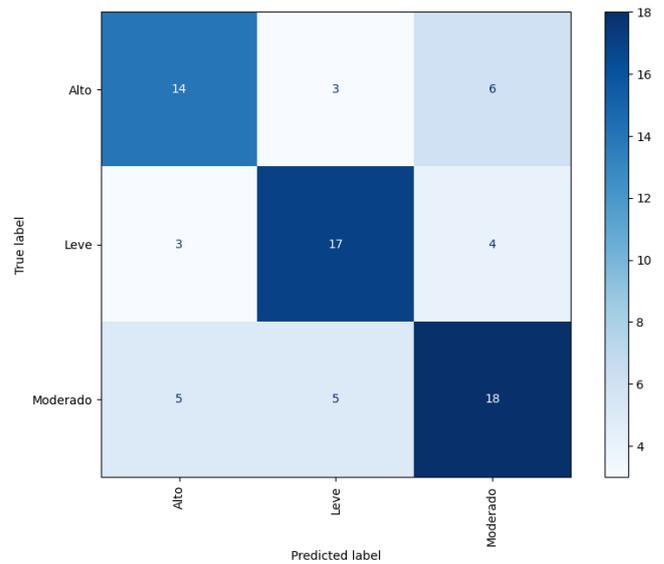


Fig. 12. Matriz de confusión Random Forest Classifier O1

El AUC del modelo de clasificación es de 0.83 para un nivel leve, 0.78 para moderado y 0.76 para un nivel alto (ver Fig. 13).

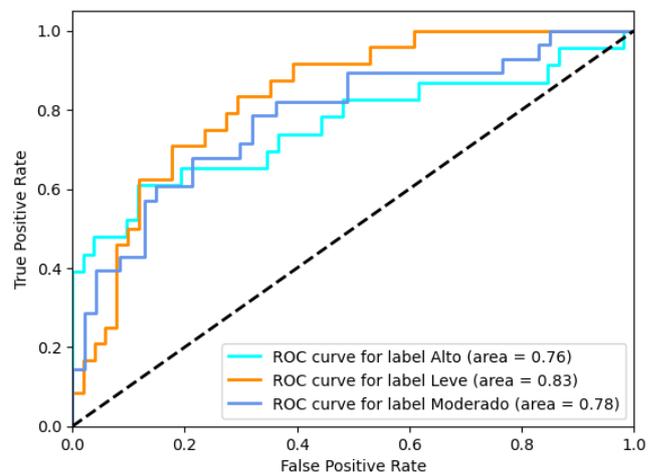


Fig. 13. Área bajo la curva ROC O1

4) Random Forest Classifier variable O2:

Para la prevalencia de inseguridad alimentaria en el hogar, se llevó a cabo el entrenamiento del algoritmo de clasificación basado en árboles de decisión Random Forest Classifier con los siguientes parámetros (ver TABLA X). Se llevo a cabo el entrenamiento con 5 pliegues para cada una de las 1,800

combinaciones de parámetros, totalizando 9,000 entrenamientos.

TABLA X. PARAMETROS PARA GRID SEARCH PARA RANDOM FOREST CLASSIFIER O2

Parámetro	Valores	Mejores parámetros
n_estimators	[50, 65, 70, 75, 80]	751
max_depth	[15, 16, 17, 18, 19]	18
min_samples_split	[2, 4, 5]	2
min_samples_leaf	[1, 2]	2
max_features	['auto', 'sqrt', 'log2']	log2
bootstrap	[False, True]	False
criterion	['gini', 'entropy']	gini
random_state	[135]	35

La matriz de confusión del modelo de clasificación mostró 30 verdaderos positivos para los niveles de prevalencia de inseguridad alimentaria leve, moderado y alto, y un total de 20 falsos positivos (ver **Fig. 14**). La accuracy del modelo fue del 60%, con una precisión de 83% para leve, 47% para moderada y 58% para alto. La medida F1-Score, que equilibra precisión y exhaustividad, alcanzó un 0.65 para leve, 0.53 para moderado y 0.63 para alto, y el Recall 0.53 (leve), 0.60 (moderado) y 0.69 (alto) (**TABLA XI**).

TABLA XI. MÉTRICAS DE DESEMPEÑO RANDOM FOREST CLASSIFIER O1

O2	Precisión	Recall	F1-Score
Alto	0.58	0.69	0.63
Leve	0.83	0.53	0.65
Moderado	0.47	0.60	0.53

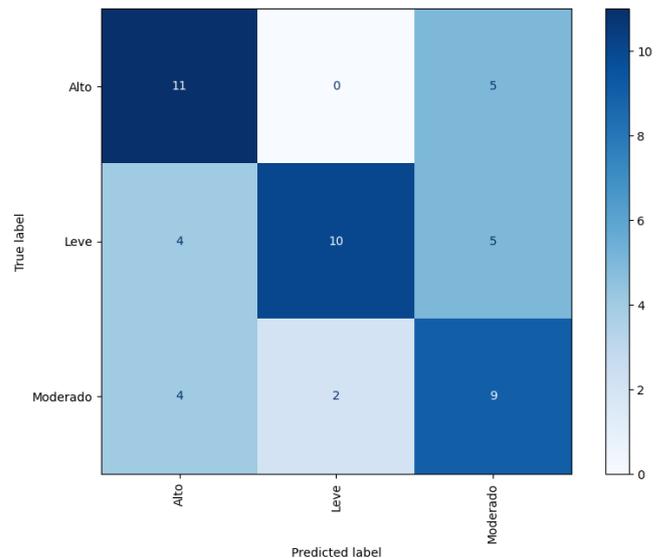


Fig. 14. Matriz de confusión Random Forest Classifier O2

El AUC del modelo de clasificación es de 0.87 para un nivel leve, 0.77 para moderado y 0.87 para un nivel alto (ver **Fig. 15**).

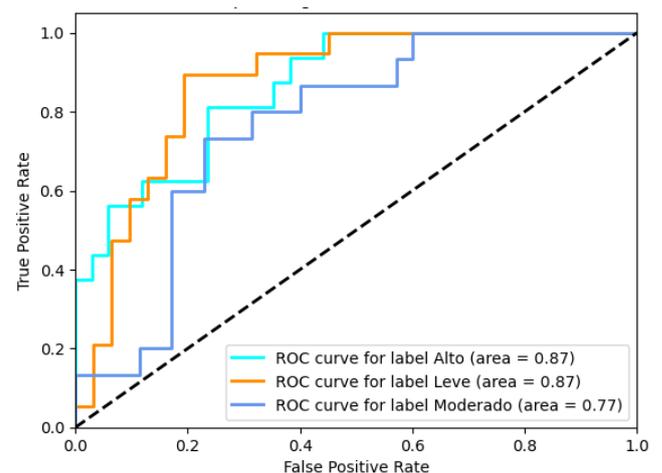


Fig. 15. Área bajo la curva ROC O2

Random Forest Classifier presentó métricas más óptimas de desempeño en los experimentos de entrenamiento realizados, siendo el mejor modelo considerado para la clasificación de la incidencia de desnutrición aguda en población menor de 5 años (Accuracy 65%, AUC 79%) y la prevalencia de inseguridad alimentaria en los hogares (Accuracy 60%, AUC 84%) del departamento de Antioquia, Colombia.

IV. CONCLUSIONES

Para el entrenamiento de los modelos se contó con una baja cantidad de datos por ser indicadores de resultado y de impacto, de variación lenta y en muchos casos representados en tasas medidas anual y bianualmente, reflejando el comportamiento del territorio a nivel municipal para el departamento de Antioquia.

La incidencia de desnutrición aguda en población menor de 5 años y la prevalencia de inseguridad alimentaria en los hogares, presentaron valores extremos, un caso particular presentó un numerador con valor de 10 y denominador poblacional con valor de 283, por lo tanto el cálculo del indicador correspondió a $\frac{10}{283} \cdot 100,000$ obteniendo una tasa que se interpreta como que cada 100,000 menores de 5 años se presentan 3,533 casos, lo cual ocasiona que la cifra aparezca bastante notable, por ejemplo en el caso de un indicador como la desnutrición.

Ya que los indicadores abordados desde el conjunto de datos en su mayoría son tasas afectadas poblacionalmente y describen el comportamiento del fenómeno en un territorio específico, no es correcto eliminar los valores que aparecen como outliers, ya que describen la criticidad del indicador en un territorio específico.

Futuras experimentaciones podrían ser abordadas con mejoras en el proceso de imputación, incluyendo métodos más rigurosos como por ejemplo la aplicación de Regresión Lineal u otras técnicas de Machine Learning y Deep Learning, además, ejecutando la imputación

variable a variable de una forma más específica. Igualmente valorando la dispersión de los datos para validar la correcta imputación y las medidas de tendencia central.

Es posible implementar mejoras en el escalamiento de las variables predictoras, mediante la exploración de resultados experimentando con diferentes métodos como el MinMaxScaler y RobustScaler, esta experimentación puede influir en el mejoramiento de las métricas para los modelos entrenados a futuro.

El coeficiente de correlación permitió priorizar 67 variables predictoras con un índice menor a 0.6, este proceso se llevó a cabo en conjunto para todas las variables y estas fueron incluidas en el entrenamiento de los modelos. A futuro es recomendable hacer un proceso de correlación para descartar o incluir variables a los modelos de forma más flexible, y considerando dichas variables de manera independiente.

La separación del dataframe permitió obtener un conjunto de datos independiente para cada variable O1 y O2, posterior se llevó a cabo la omisión de registros con valores nulos en dichos indicadores para finalmente proceder con el entrenamiento de los modelos. El mejor modelo se obtuvo con la variable O1, posiblemente por contener mayor cantidad de instancias que pudieron influir en los resultados obtenidos.

Experimentar otro tipo de modelos como Redes Neuronales puede facilitar la obtención de mejores métricas. Igualmente, los resultados obtenidos en el entrenamiento de los modelos

evidenciaron métricas sobresalientes para Extra Trees Classifier y Random Forest Classifier.

Estos resultados indican un rendimiento generalmente aceptable del modelo, lo que revela que se posee una buena capacidad para distinguir entre clases positivas y negativas. En promedio, Extra Trees Classifier obtuvo un 80% de probabilidad de clasificar correctamente los niveles de desnutrición y de inseguridad alimentaria, mientras que Random Forest Classifier evidencio 79% y 84% para O1 y O2 respectivamente.

Estos resultados indican un rendimiento generalmente aceptable del modelo, el cual puede mejorar en la identificación de los casos positivos.

Con respecto a la usabilidad de los modelos finales, estos permiten identificar para los municipios las variables críticas para abordar y a su vez, apoyar el modelamiento de respuestas ante posibles cambios en las condiciones de los territorios, como predecir el impacto de ciertas políticas públicas.

Es posible usar estos modelos, por su rendimiento aceptable, como punto de partida para proyectar alcances más ambiciosos en la clasificación de la desnutrición y la inseguridad alimentaria.

V. REFERENCIAS

- [1] Organización Mundial de la Salud, “Nutrición,” 2024. [Online]. Available: <https://www.who.int/es/health-topics/nutrition>
- [2] Organización Mundial de la Salud, “Malnutrición,” 2024. [Online]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/malnutrition>
- [3] Organización de las Naciones Unidas para la Alimentación y la Agricultura, Fondo Internacional de Desarrollo Agrícola, Fondo de las Naciones Unidad para la Infancia, Programa Mundial de Alimentos, and Organización Panamericana de la Salud, *Panorama Regional De Seguridad Alimentaria Y Nutricional*. 2023. [Online]. Available: <https://iris.paho.org/handle/10665.2/57048>
- [4] N. Unidas, “Informe de los Objetivos de Desarrollo Sostenible,” 2023. [Online]. Available: https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023_Spanish.pdf?_gl=1*1vtezeu*_ga*MjEyODc3Nzk3Mi4xNjk1MzQ5NzQ5*_ga_TK9BQL5X7Z*MTY5NTM0OTc0OS4xLjEuMTY5NTM1MDA4Ny4wLjAuMA
- [5] N. C. R. V. Monica Paola Castro Soto, Luisa Fernanda Echeverri Buriticá, Luisa Fernanda Londoño Giraldo, Diana Sorley Muñoz Mazo, “Mortalidad en niños menores de cinco años por desnutrición, en la comunidad indígena Arara de Leticia, Amazonas,” pp. 1–29, 2024, [Online]. Available: <https://repository.unad.edu.co/bitstream/handle/10596/62113/lflondonogi.pdf?sequence=3&isAllowed=y>
- [6] G. Congacha, “Comparación de modelos logísticos y árboles de decisión para

- identificar y predecir factores asociados a la desnutrición crónica infantil basados en la encuesta nacional de salud y nutrición – ENSANUT 2018-2019,” *Esc. Super. Politécnica Chimborazo*, pp. 1–67, 2020, [Online]. Available: [http://dspace.esPOCH.edu.ec/bitstream/123456789/14551/1/226T0063 %281%29.pdf](http://dspace.esPOCH.edu.ec/bitstream/123456789/14551/1/226T0063%281%29.pdf)
- [7] C. Machaca, “Modelo Univariante Para La Predicción De La Desnutrición Crónica De Los Niños Menores De Cinco Años En El Hospital Regional Manuel Nuñez Butrón,Puno 2012 - 2016.,” p. 98, 2017, [Online]. Available: http://repositorio.unap.edu.pe/bitstream/handle/UNAP/9386/Machaca_Uriel_Clever_Uriel.pdf?sequence=3&isAllowed=y
- [8] O. Reséndiz, C. Torres, A. Arredondo, Y. García, and E. Orozco, “Determinantes de la inseguridad alimentaria y su asociación con prácticas de alimentación y lactancia materna en el primer año de vida,” *Glob. Health Promot.*, vol. 29, no. 4, pp. 161–170, 2021, doi: 10.1177/17579759221076881.
- [9] Gobernación de Antioquia and Universidad de Antioquia, “Perfil Alimentario y Nutricional de Antioquia 2019,” *Perf. Aliemntario y Nutr. Antioquia 2019*, p. 70, 2019, [Online]. Available: https://www.udea.edu.co/wps/wcm/connect/udea/ba8c4eac-c8ad-43aa-9a6f-7f9f143d28a0/resumen_ejecutivo_perfil_alimentario_FINAL.pdf?MOD=AJPERES&CVID=n3658bG
- [10] Á. A. Loaiza Duque, W. G. Moreno Puerta, and J. D. Rios Rios, “Proceso de analítica de datos aplicado a la desnutrición infantil en niños de 0 a 5 años en la ciudad de Medellín,” 2023, [Online]. Available: http://siawebpr.itm.edu.co/bitstream/handle/20.500.12622/6011/AlvaroAndres_LoalizaDuque_WilliamGuillermo_MorenoPuerta_JuanDavid_RiosRios_2023.pdf?sequence=2&isAllowed=yes
- [11] A. Ávila Curiel *et al.*, “Mala nutrición en población escolar mexicana: factores geográficos y escolares asociados,” *Glob. Health Promot.*, vol. 29, no. 2, pp. 126–135, 2021, doi: 10.1177/17579759211038381.
- [12] H. F. Martínez, “Desnutrición a nivel municipal en Colombia,” *República Colomb. Nac. Planeación-Dirección Estud. Económicos-Archivos Econ.*, pp. 1–34, 2009, [Online]. Available: <https://www.dnp.gov.co/LinkClick.aspx?fileticket=cbZhW6NhfZE%3D&tabid=897>
- [13] G. Hoyos, Y. Beatriz, A. Dra, M. Colmenares, and P. Clementina, “Determinantes socioeconómicos de la desnutrición crónica infantil en el Perú , año 2021,” 2023.
- [14] A. Vilca Mamani, E. B. García Castro, L. Lipa Tudela, E. Calancho Mamani, and R. M. Cruz Huisa, “Impacto de los programas sociales alimentarios sobre la desnutrición infantil en la región de Puno,” *Comuni@cción Rev. Investig. en Comun. y Desarro.*, vol. 14, no. 3, pp. 220–234, 2023,

- doi: 10.33595/2226-1478.14.3.871.
- [15] G. Estrada and L. Roldán, “Modelo estadístico para predecir la prevalencia de desnutrición crónica infantil en los departamentos de Guatemala,” *Rev. médica (Colegio Médicos y Cir. Guatemala)*, vol. 156, no. 2, pp. 61–66, 2017, doi: 10.36109/rmg.v156i2.57.
- [16] P. Flores and G. Congacha, “Factores Asociados a La Desnutrición Crónica Infantil En Ecuador. Estudio Basado En Modelos De Regresión Y Árboles De Clasificación.,” *Perfiles*, vol. 1, no. 26, pp. 21–33, 2021, doi: 10.47187/perf.v1i26.132.
- [17] F. DE Mecánica, P. Por, and J. Diego Cruz Freire Darwin Vinicio Chimbo Chimbo, “Factores asociados en la desnutrición crónica en niños menores de 5 años en el Hospital Alfonso Villagómez, mediante la comparación de modelo de regresión y análisis de correspondencia en el periodo 2020-2022,” 2015.
- [18] Norabuena Figueroa Roger Pedro, *Random forest de los factores asociados a la desnutrición crónica en niños peruanos menores de 5 años 2020*. 2023. [Online]. Available: https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/19783/Paucar_gb.pdf?sequence=1&isAllowed=y
- [19] C. Stix, “Actionable Principles for Artificial Intelligence Policy: Three Pathways,” *Sci. Eng. Ethics*, vol. 27, no. 1, pp. 1–17, 2021, doi: 10.1007/s11948-020-00277-3.
- [20] F. H. Arnold, M. T. Zuber, and Prabhakar Arati, “Supercharging Research: Harnessing Artificial Intelligence to Meet Global Challenges,” no. April, 2024, [Online]. Available: www.whitehouse.gov/pcast.
- [21] Huberth Rolando Hincapié Arango. Yeisson Alejandro Gutiérrez Cano, “Monografía-Huberth-Yeisson,” GitHub. [Online]. Available: <https://github.com/AlejandroGutie/Monografia-Huberth-Yeisson.git>
- [22] Secretaría Seccional de Salud de Antioquia, “Servicio de Información y Estadística,” Gobernación de Antioquia. [Online]. Available: <https://dssa.gov.co/inicio-estadisticas>
- [23] Gobernación de Antioquia, “Observatorio de Políticas Públicas de Antioquia,” www.antioquiadatos.gov.co. [Online]. Available: <https://www.antioquiadatos.gov.co/index.php/oppa/>
- [24] Secretaría Seccional de Salud de Antioquia, “Agenda Antioquia 2040,” Gobernación de Antioquia. [Online]. Available: <https://agendaantioquia.com/biblioteca/>
- [25] Secretaría Seccional de Salud de Antioquia, “Estadística Vitales,” Gobernación de Antioquia. [Online]. Available: <https://dssa.gov.co/estadisticas-vitales>
- [26] pandas.pydata.org, “pandas.pivot_table,” OVHcloud. [Online]. Available: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.pivot_table

- e.html
- [27] K_DM, “Coeficiente de correlación,” k-dm.work. [Online]. Available: https://k-dm.work/es/eval/regression/correlation_coefficient/
- [28] Python Software Foundation, “statistics — Funciones de estadística matemática,” docs.python.org. [Online]. Available: <https://docs.python.org/es/3.10/library/statistics.html>
- [29] J. D. Hernández-Lalinde *et al.*, “Sobre el uso adecuado del coeficiente de correlación de Pearson,” *Arch. Venez. Farmacol. y Ter.*, vol. 37, no. 5, pp. 586–601, 2019, [Online]. Available: http://saber.ucv.ve/ojs/index.php/rev_aavft/article/view/16165
- [30] scikit-learn developers, “Scikit Learn, KNNImputer (k-Nearest Neighbors),” scikit-learn.org. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html>
- [31] F. I. Araneda, Rodrigo Ignacio Navarro. Dra, Rosa L, “Algoritmo de aprendizaje automático para el estudio de la asociación entre enfermedad cardiovascular y depresión,” pp. 356–363, 2023, [Online]. Available: http://repositorio.udec.cl/jspui/bitstream/11594/11348/1/Navarro_Araneda_Rodrigo_Tesis.pdf
- [32] N. Emiliano García D’Urso, A. Fuster Guilló, and J. Azorín López, “Análisis de la evolución morfológica del cuerpo y los datos clínicos de pacientes en tratamiento dietético nutricional,” 2021, [Online]. Available: https://rua.ua.es/dspace/bitstream/10045/118325/1/Analisis_de_la_evolucion_morfolo_gica_del_cuerpo_Garcia_DUrso_Nahuel_Emiliano.pdf
- [33] Pandas.pydata.org, “Pandas.qcut,” OVHcloud. [Online]. Available: <https://pandas.pydata.org/docs/reference/api/pandas.qcut.html>
- [34] M. Ali, “PyCaret, low-code machine learning,” pycaret.gitbook.io. [Online]. Available: <https://pycaret.org/>
- [35] scikit-learn developers, “Tuning the hyperparameters of an estimator,” scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/grid_search.html#grid-search
- [36] R. Borja-Robalino, A. Monleon-Getino, and J. Rodellar, “Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning,” *Rev. Ibérica Sist. e Tecnol. Inf.*, vol. 30, no. June, pp. 184–196, 2020, [Online]. Available: <https://www.researchgate.net/publication/342009715>
- [37] V. D. Gil-Vera and C. Quintero-López, “Predicción del rendimiento académico estudiantil con redes neuronales artificiales,” *Inf. tecnológica*, vol. 32, no. 6, pp. 221–228, 2021, doi: 10.4067/s0718-07642021000600221.

- [38] J. J. Saldana-Barrios, T. Concepción, and M. Vargas-Lombardo, “Mejorando la predicción del síndrome de Down mediante un modelo de clasificación de datos médicos inteligente- Caso de Estudio,” *I+D Tecnológico*, vol. 12, no. 2, pp. 36–45, 2016, [Online]. Available: <http://revistas.utp.ac.pa/index.php/id-tecnologico/article/view/1234>
- [39] scikit-learn developers, “KFold,” scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html
- [40] scikit-learn developers, “Cross-validation: evaluating estimator performance,” scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html