

<b>Facultad de Ciencias Naturales y Matemáticas</b>	<b>FORMATO DE ANTEPROYECTO</b>	
	<b>CÓDIGO:</b>	<b>APROBADO POR:</b> <b>Facultad de Ciencias Naturales y Matemáticas</b>
	<b>VERSIÓN:</b>	
	<b>FECHA:</b>	

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**  
**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS**  
**UNIDAD DE POSTGRADO**  
***Proyecto de Investigación para Titulación***  
**MAESTRÍA EN ESTADÍSTICA APLICADA**

**Acerca del Estudiante:**

Nombre del estudiante:	Ing. Marcelo Raúl Chávez Reinoso
No. cédula de ciudadanía:	0603065186
Correo electrónico:	marcrein@espol.edu.ec

**Acerca del Tutor del Proyecto:**

Nombre del Tutor:	
Nivel de instrucción más alto:	PhD. Sergio Bauz Olvera
Áreas de especialidad:	Machine Learning y Métodos Multivariantes
Correo electrónico:	serabauz@espol.edu.ec

**Acerca del Proyecto:**

<b>Título del Proyecto de Investigación:</b>
Modelos de ensamble en Machine Learning para la identificación de los factores relevantes en la Desnutrición Crónica Infantil y la zonificación territorial a través de un Índice Multivariante automático.
<b>Planteamiento del Problema:</b>
<p>En Ecuador 1 de cada 4 niños y niñas menores de 5 años presenta desnutrición crónica infantil<sup>1</sup>, esto nos alerta como país para identificar la verdadera realidad de su origen y formas de erradicación. La desnutrición crónica infantil afecta al 27,2% de los niños menores de 2 años en Ecuador. (Resultados de ENSANUT 2018 – INEC). Las causas de la malnutrición -desnutrición y sobrepeso– que se agrupan en factores causales de tipo inmediato, subyacentes y básicos (Rivera Jairo, 2022). Es una problemática que repercute en la productividad de un país y tiene un impacto a lo largo de la vida de las personas, y puede provenir de un proceso sistémico para las condiciones de progreso de un gobierno de forma estructural.</p> <p>También resultante importante analizar que, en el sector indígena, según la Encuesta Nacional de Salud y Nutrición (ENSANUT 2018) 4 de cada 10 niños sufren de DCI, esto no solo porque es un proceso intrínseco de este sector de la población, sino que es generado por determinantes sociales y factores subyacentes como el consumo inadecuado de alimentos, acceso deficiente a servicios básicos, prácticas inadecuadas de atención médica y cuidado materno-infantil, entre otros.</p> <p>Es así que es imperante desarrollar una técnica robusta de predicción de las condiciones o factores asociados a la Desnutrición Infantil con nuevos enfoques de aprendizaje estadístico (Public Health Nutrition, 2022), para lo cual en el presente proyecto de</p>

<sup>1</sup> Fuente: Instituto Nacional de Estadística y Censos (INEC) – ENSANUT 2018

investigación se implementará una Metodología de Modelos de Ensamble y un Método Multivariante que permita clasificar y predecir las condiciones de Desnutrición a menores de 5 años, tomando como población objetivo la investigada en el 2022 en la Encuesta Nacional de Desnutrición Infantil, por el Instituto Nacional de Estadística y Censos (INEC).

### Justificación:

La desnutrición crónica infantil es un problema de salud pública de gran magnitud en Ecuador, afectando a una proporción significativa de la población infantil, especialmente en comunidades vulnerables como las indígenas. Esta condición no solo tiene consecuencias inmediatas en la salud y el bienestar de los niños, sino que también tiene repercusiones a largo plazo en su desarrollo físico, cognitivo y socioemocional. Además, la desnutrición infantil está estrechamente vinculada con la pobreza, la falta de acceso a servicios básicos de salud y nutrición, y otros determinantes sociales que perpetúan el ciclo de la pobreza y la desigualdad en el país (Rivera Jairo, 2022).

En este contexto, es imperativo abordar la desnutrición crónica infantil de manera integral y efectiva. Sin embargo, para diseñar políticas y programas de intervención adecuados, es fundamental comprender en profundidad los factores subyacentes que contribuyen a esta problemática y cómo estos factores interactúan entre sí en diferentes contextos sociodemográficos y geográficos.


Es aquí donde radica la importancia de desarrollar una metodología robusta de predicción de las condiciones o factores asociados a la desnutrición infantil. Utilizando enfoques innovadores de aprendizaje estadístico, como modelos de ensamble y métodos multivariantes, podemos analizar grandes volúmenes de datos para identificar patrones, tendencias y relaciones complejas que pueden no ser evidentes mediante métodos tradicionales de análisis estadístico.

Al implementar esta metodología en el presente proyecto de investigación y utilizando datos recientes de la Encuesta Nacional de Desnutrición Infantil realizada por el INEC, se espera generar información precisa y relevante sobre los determinantes de la desnutrición crónica infantil en Ecuador. Esto no solo permitirá una mejor comprensión de la problemática, sino que también sentará las bases para el diseño y la implementación de intervenciones más efectivas y orientadas a la prevención y el tratamiento de la desnutrición infantil en el país.

En última instancia, este estudio tiene el potencial de contribuir significativamente a la mejora de la salud y el bienestar de los niños ecuatorianos, así como a la reducción de las desigualdades sociales y la promoción de un desarrollo sostenible y equitativo en el país.

### Objetivo General:

Desarrollar una metodología de modelos de ensamblaje asociados a un método multivariante para determinar los principales factores más importantes en la desnutrición crónica infantil en niños menores de 5 años en Ecuador, utilizando como

 <b>Facultad de Ciencias Naturales y Matemáticas</b>	<b>INFORME DE EVALUACIÓN DEL TRABAJO DE TITULACIÓN PREVIO A LA DEFENSA PÚBLICA</b>	
	<b>CÓDIGO:</b> FT-21-V01	<b>APROBADO POR:</b> <b>Facultad de Ciencias Naturales y Matemáticas</b>
	<b>VERSIÓN:</b> 01	
	<b>FECHA:</b> 05/10/2021	

fuentes oficiales la Encuesta Nacional de Desnutrición Infantil del Instituto Nacional de Estadística y Censos (INEC) del año 2022

#### **Objetivos Específicos:**

- ✓ Estudiar los principales modelos multivariantes y de ensamblaje a través de una revisión bibliográfica exhaustiva sobre el contexto de la desnutrición infantil.
- ✓ Implementar modelos de ensamblaje asociados a un método multivariante para clasificar los factores asociados a la desnutrición crónica infantil con base a variables demográficas, socioeconómicas, antropométricas y de acceso a servicios básicos de salud y nutrición.
- ✓ Evaluar los resultados de los modelos de ensamble y multivariante para las relaciones de las variables predictoras de la desnutrición crónica infantil, permitiendo una comprensión territorial de los patrones subyacentes a través de una visualización espacial en las distintas poblaciones y regiones del Ecuador.

#### **Hipótesis (Opcional, depende del tipo de proyecto):**

La hipótesis de investigación es determinar si las relaciones causales entre variables demográficas, socioeconómicas, antropométricas, acceso a servicios básicos de salud y nutrición son determinantes para identificar y clasificar la Desnutrición Crónica Infantil, según la Encuesta Nacional de Desnutrición Infantil del Instituto Nacional de Estadística y Censos (INEC) del año 2022.

#### **Alcance:**

El alcance del proyecto se enfoca en la aplicación de técnicas estadísticas avanzadas para abordar el desafío de la desnutrición crónica infantil en Ecuador, una problemática de salud pública que afecta significativamente a la población infantil, especialmente en comunidades vulnerables (Rivera Jairo, 2022). Para ello, se llevará a cabo un exhaustivo análisis de datos provenientes de la Encuesta Nacional de Desnutrición Infantil del Instituto Nacional de Estadística y Censos (INEC) del año 2022, así como de otras fuentes pertinentes de información demográfica, socioeconómica y de salud.

En primer lugar, se realizará una fase de preparación y limpieza de datos, donde se integrarán y depurarán los conjuntos de datos disponibles para su posterior análisis. Posteriormente, se aplicarán modelos de ensamble, como Random Forest y Gradient Boosting, para la predicción de la desnutrición crónica infantil. Estos modelos permitirán identificar patrones complejos y relaciones no lineales entre las variables predictoras y la presencia de desnutrición en los niños menores de 5 años.

Además, se utilizará un método multivariante, para explorar la estructura subyacente de los datos y entender mejor los factores que contribuyen a la desnutrición crónica infantil en diferentes contextos. Esta etapa será fundamental para identificar variables clave y comprender cómo interactúan entre sí para influir en la prevalencia de la desnutrición en distintas poblaciones y regiones del país.

Una vez desarrollados los modelos predictivos y analizados los resultados obtenidos, se procederá a la zonificación territorial. Esta fase implicará la elaboración de mapas que identifiquen las áreas con mayor riesgo de desnutrición infantil, lo que permitirá a las

autoridades y organizaciones pertinentes priorizar intervenciones y asignar recursos de manera más eficiente y efectiva.

Finalmente, se llevará a cabo la validación de los modelos y la comunicación de los resultados a través de informes técnicos y presentaciones destinadas a tomadores de decisiones en el ámbito gubernamental, organizaciones no gubernamentales y otros actores relevantes. El objetivo es proporcionar evidencia sólida y basada en datos para informar y respaldar la formulación de políticas y programas dirigidos a la prevención y tratamiento de la desnutrición crónica infantil en Ecuador.

### Metodología:

La metodología propuesta para abordar la predicción de la desnutrición crónica infantil y la zonificación territorial en Ecuador se enfocará en el desarrollo de un proceso estadístico con un enfoque de Machine Learning:

1. **Recopilación de datos:** Estructuración de la base de datos de trabajo a partir de la Encuesta Nacional de Desnutrición Infantil del INEC del año 2022, así como otras fuentes de datos relevantes sobre variables demográficas, socioeconómicas y de acceso a servicios de salud y nutrición. Estos datos se organizarán y prepararán para su posterior análisis.
2. **Preprocesamiento de datos:** Fase de limpieza y transformación de datos para abordar problemas como valores atípicos, datos faltantes y errores de entrada. Además, se realizará una exploración inicial de los datos para comprender su distribución y características.
3. **Selección de factores subyacentes o constructos:** Con base a fundamentos teóricos sobre investigaciones desarrolladas sobre la Desnutrición Crónica Infantil se identificarán las variables más relevantes que podrían estar asociadas con la desnutrición crónica infantil. Esto podría incluir métodos estadísticos como la correlación, así como técnicas de asociación de variables, entre otros.

### Factores o constructos que se han definido a priori para la presente investigación:

- ✓ Variables antropométricas
  - ✓ Variables de nutrición
  - ✓ Variables de acceso a servicios básicos
  - ✓ Variables de salud del niño (menor de 5 años) o gestante.
4. **Desarrollo de modelos de ensamble:** Se implementarán modelos de ensamble, como Random Forest y Gradient Boosting, utilizando bibliotecas de aprendizaje automático como Tidymodels (librería del Software **R**). Estos modelos se entrenarán utilizando los datos preparados y se optimizarán mediante técnicas como la búsqueda de hiperparámetros.
  5. **Aplicación de métodos multivariantes:** Se utilizarán métodos multivariantes, como el análisis de Correspondencias Múltiples y el Análisis Discriminante, para explorar la estructura subyacente de los datos y reducir la dimensionalidad

cuando sea necesario. Esto permitirá una comprensión más profunda de las relaciones entre las variables predictoras y la desnutrición crónica infantil.


6. **Validación del modelo:** Se realizará una validación cruzada y otros métodos de validación para evaluar la precisión y la generalización de los modelos desarrollados. Esto ayudará a garantizar que los modelos sean robustos y puedan aplicarse a datos nuevos, a través de la técnica K-Folds.
7. **Zonificación territorial:** Utilizando los resultados del método multivariante, se procederá a la zonificación territorial para identificar áreas con diferentes niveles de riesgo de desnutrición crónica infantil. Y su posterior visualización para representar los resultados de manera clara y comprensible.
8. **Visualización de datos:** Se elaborará un **aplicativo web** en el **Framework Shiny** del Software **R** para mostrar los resultados de manera clara y accesible para los tomadores de decisiones y la comunidad en general, y se proporcionarán recomendaciones basadas en evidencia para informar sobre los factores latentes que influyen en la identificación de la Desnutrición en menores de 5 años, y que permitan la formulación de nueva política pública orientada a la prevención e identificación de la desnutrición crónica infantil en Ecuador.

#### Datos:

- Encuesta Nacional de Desnutrición Infantil del Instituto Nacional de Estadística y Censos (INEC) del año 2022.
- Base de datos del Censo de Población y Vivienda, INEC 2022.

#### Revisión Bibliográfica (Estado del Arte):

- Rivera Vásquez, Jairo Israel. 2020. La malnutrición infantil en Ecuador: entre progresos y desafíos. Tesis de doctorado, FLACSO Ecuador.
- Johnson, R. A., & Wichern, D. W. (2019). Applied Multivariate Statistical Analysis (6th ed.). Pearson.
- Elsevier. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. Nutrition, 78, 1108610.
- Public Health Nutrition. (2022). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. Public Health Nutrition, 25(2), 269–280
- Public Health Nutrition. (2022). A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. Public Health Nutrition
- International Journal of Cognitive Computing in Engineering. (2022). Application of machine learning-based algorithm for prediction of malnutrition among women in Bangladesh. International Journal of Cognitive Computing in Engineering.
- Research Square. (2023, 9 de junio). Prediction of malnutrition in newborn infants using machine learning techniques.
- Indian Institute of Technology Jodhpur, India, & All India Institute of Medical Science Jodhpur, India. NutriAI: AI-powered child malnutrition assessment in low-resource environments.

 <b>Facultad de Ciencias Naturales y Matemáticas</b>	<b>INFORME DE EVALUACIÓN DEL TRABAJO DE TITULACIÓN PREVIO A LA DEFENSA PÚBLICA</b>	
	<b>CÓDIGO:</b> FT-21-V01	<b>APROBADO POR:</b> <b>Facultad de Ciencias Naturales y Matemáticas</b>
	<b>VERSIÓN:</b> 01	
	<b>FECHA:</b> 05/10/2021	

- School of Public Health, Chongqing Medical University, Chongqing 400016, China. Machine learning algorithms for predicting stunting among under-five children in Papua New Guinea.
- Kar, S., Pratihari, S., Nayak, S., Bal, S., & Gururaj, H. L. (2021, diciembre). Prediction of child malnutrition using machine learning. IEEE Xplore.
- Sharma, V., Sharma, V., Khan, A., Wassmer, D. J., Schoenholtz, M. D., Hontecillas, R., ... Abedi, V. (2020, 15 de abril). Nutrition and Immunology. Frontiers in Nutrition.

#### **Resultados Esperados:**

- Metodología de ensamblado y multivariante para la caracterización de los factores asociados a la desnutrición crónica infantil
- Zonificación territorial sobre las poblaciones de niños menores de 5 años con desnutrición en Ecuador
- Priorización de áreas o variables de estudio en el contexto de la desnutrición infantil

#### **Bibliografía:**

- Rivera Vásquez, Jairo Israel. 2020. La malnutrición infantil en Ecuador: entre progresos y desafíos. Tesis de doctorado, FLACSO Ecuador.
- Johnson, R. A., & Wichern, D. W. (2019). Applied Multivariate Statistical Analysis (6th ed.). Pearson.
- Elsevier. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. Nutrition, 78, 1108610.
- Public Health Nutrition. (2022). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. Public Health Nutrition, 25(2), 269–280
- Public Health Nutrition. (2022). A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. Public Health Nutrition
- International Journal of Cognitive Computing in Engineering. (2022). Application of machine learning-based algorithm for prediction of malnutrition among women in Bangladesh. International Journal of Cognitive Computing in Engineering.
- Research Square. (2023, 9 de junio). Prediction of malnutrition in newborn infants using machine learning techniques.
- Indian Institute of Technology Jodhpur, India, & All India Institute of Medical Science Jodhpur, India. NutriAI: AI-powered child malnutrition assessment in low-resource environments.
- School of Public Health, Chongqing Medical University, Chongqing 400016, China. Machine learning algorithms for predicting stunting among under-five children in Papua New Guinea.
- Kar, S., Pratihari, S., Nayak, S., Bal, S., & Gururaj, H. L. (2021, diciembre). Prediction of child malnutrition using machine learning. IEEE Xplore.



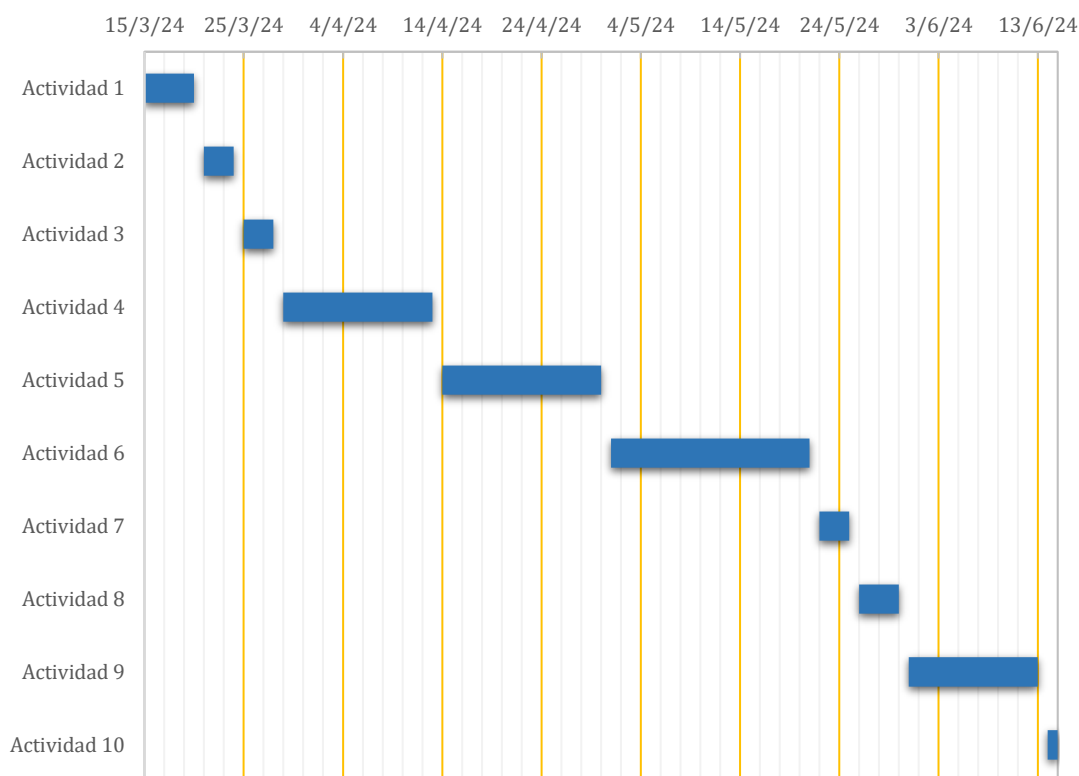
- Sharma, V., Sharma, V., Khan, A., Wassmer, D. J., Schoenholtz, M. D., Hontecillas, R., ... Abedi, V. (2020, 15 de abril). Nutrition and Immunology. Frontiers in Nutrition.
- Hastie, T., Tibshirani, R., & Friedman, J. (Año). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.

### Cronograma:

El proyecto de investigación contiene los siguientes hitos más relevantes para su desarrollo:

Nro.	Actividades	Fecha de inicio	Duración en días	Fecha fin
Actividad 1	Revisión literaria sobre investigaciones sobre la Desnutrición Infantil y recopilación de la bibliografía técnica sobre los métodos de ensamble	15/3/2024	5	20/3/2024
Actividad 2	Estructuración de las bases de datos para trabajo sobre la Encuesta ENDI y el Censo de Población como fuentes de información oficial para el desarrollo de factores y constructos.	21/3/2024	3	24/3/2024
Actividad 3	Desarrollo del Análisis Exploratorio de la población objetivo utilizando los datos recopilados.	25/3/2024	3	28/3/2024
Actividad 4	Desarrollo de la Introducción, antecedentes, hipótesis, objetivos, y marco teórico	29/3/2024	15	13/4/2024
Actividad 5	Desarrollo del objeto de investigación, modelos de ensamble, evaluación de los modelos y ajustes	14/4/2024	16	30/4/2024
Actividad 6	Desarrollo de un aplicativo en el Framework Shiny de R para automatización y visualización de los resultados	1/5/2024	20	21/5/2024
Actividad 7	Redacción de conclusiones finales y preparación del informe final	22/5/2024	3	25/5/2024
Actividad 8	Presentación del informe final, junto con la entrega del aplicativo desarrollado	26/5/2024	4	30/5/2024
Actividad 9	Revisión, cambios, ajustes y observaciones por parte del Tutor de la Investigación y el tesista	31/5/2024	14	13/6/2024
Actividad 10	Entrega del Proyecto de Investigación	14/6/2024	1	15/6/2024

### Diagrama de Gantt



**Elaborado por:**

Ing. Marcelo Raúl Chávez Reinoso

**Tutoriado por:**

PhD. Sergio Bauz Olvera

**Fecha de Recepción:**  
26 de febrero de 2024

**Fecha de Aprobación:**  
26 de febrero de 2024