

Desigualdades en atención sanitaria: Ciencia de Datos y Machine Learning

Edwin Molano Suárez

Asesor

Luis Angel Anillo Arrieta

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Sociales Artes y Humanidades ECSAH

Especialización en Ciencia de Datos y Analítica

2025

Nota de Aceptación

Nombre Director de Trabajo de Grado

Jurado

Jurado

Resumen

La propuesta de monografía titulada “Desigualdades en Atención Sanitaria: Ciencia de Datos y Machine Learning” se centra en analizar las brechas existentes en el acceso a los servicios de salud, aplicando herramientas tecnológicas de analítica de datos y aprendizaje automático para identificar patrones de inequidad y proponer soluciones basadas en evidencia.

En el resumen se expone que las desigualdades sanitarias constituyen disparidades evitables en el acceso y los resultados de la atención médica, afectando principalmente a poblaciones vulnerables por factores socioeconómicos, geográficos y estructurales. Se plantea que el uso de la ciencia de datos permite identificar tendencias y mejorar la toma de decisiones, orientando políticas más equitativas.

La justificación resalta que estas desigualdades representan no solo un problema de salud pública, sino también una violación de derechos fundamentales. Abordarlas es clave para promover cohesión social, equidad y sostenibilidad en los sistemas de salud. Se subraya la falta de datos desagregados y de enfoques tecnológicos innovadores, lo cual limita las intervenciones actuales. Por ello, el proyecto busca aportar desde la ingeniería y la ciencia de datos, complementando estudios cualitativos con modelos predictivos que generen recomendaciones aplicables a políticas públicas.

El objetivo general consiste en identificar variables determinantes de las desigualdades sanitarias mediante análisis de datos, detección de patrones y formulación de propuestas para mejorar el acceso a servicios de salud. Los objetivos específicos incluyen: analizar factores socioeconómicos, demográficos y geográficos que condicionan la inequidad, desarrollar un modelo predictivo que permita localizar zonas críticas y grupos afectados, y diseñar estrategias de optimización de recursos basadas en datos.

El marco conceptual y teórico se estructura en torno a tres ejes: los determinantes sociales de la salud, que evidencian la influencia del nivel socioeconómico, la educación, la discriminación y la localización geográfica en el acceso a la atención; la discriminación estructural, que afecta a grupos minoritarios como indígenas, migrantes y comunidades rurales; y el papel de la tecnología en salud, destacando la telemedicina y los sistemas digitales, cuyo acceso desigual puede profundizar las brechas existentes. Se apoya en teorías de salud pública y en estudios recientes sobre la aplicación del big data y el machine learning en la gestión de inequidades sanitarias.

En cuanto a la metodología, se empleará un enfoque cuantitativo basado en técnicas de ciencia de datos, minería de información y algoritmos de aprendizaje automático, con el fin de identificar correlaciones y tendencias ocultas en grandes volúmenes de información sanitaria. Asimismo, se contempla el uso del método PRISMA para la revisión sistemática de literatura, garantizando rigor científico en la selección de fuentes y estudios previos.

El cronograma proyecta un periodo de ejecución, dividido en fases que van desde la recolección de datos, análisis descriptivo, desarrollo de estrategias, implementación de un piloto en una región seleccionada, hasta la evaluación del impacto y la difusión de resultados.

En cuanto a recursos, se prevé el uso de software de análisis estadístico y de machine learning, acceso a bases de datos sanitarias, infraestructura tecnológica segura para el manejo de información sensible, y apoyo del asesor académico.

Los resultados esperados incluyen la identificación de factores críticos de inequidad, la validación de un modelo predictivo para focalizar recursos en áreas vulnerables, y la elaboración de recomendaciones aplicables a políticas públicas. Asimismo, se espera generar publicaciones científicas y presentaciones en escenarios académicos y de toma de decisiones.

Finalmente, el proyecto contó con el acompañamiento de la asesora Nidia Danigza Lugo López, docente, quien guio la ejecución metodológica y la rigurosidad académica de la investigación.

Palabras clave: Desigualdad; Salud; Vulnerabilidad; Atención; Análisis.

Abstract

The proposed monograph, titled “Healthcare Inequalities: Data Science and Machine Learning,” focuses on analyzing existing gaps in access to health services, applying data analytics and machine learning technologies to identify patterns of inequity and propose evidence-based solutions.

The abstract explains that health inequalities constitute avoidable disparities in access to and outcomes of medical care, primarily affecting vulnerable populations due to socioeconomic, geographic, and structural factors. It argues that the use of data science allows for the identification of trends and improved decision-making, guiding more equitable policies.

The rationale highlights that these inequalities represent not only a public health problem but also a violation of fundamental rights. Addressing them is key to promoting social cohesion, equity, and sustainability in health systems. The lack of disaggregated data and innovative technological approaches is underscored, as this limits current interventions. Therefore, the project aims to contribute from the perspectives of engineering and data science, complementing qualitative studies with predictive models that generate recommendations applicable to public policy.

The overall objective is to identify variables that determine health inequalities through data analysis, pattern detection, and the formulation of proposals to improve access to health services. Specific objectives include: analyzing socioeconomic, demographic, and geographic factors that contribute to inequity; developing a predictive model to identify critical areas and affected groups; and designing data-driven resource optimization strategies.

The conceptual and theoretical framework is structured around three main areas: the social determinants of health, which demonstrate the influence of socioeconomic status,

education, discrimination, and geographic location on access to care; structural discrimination, which affects minority groups such as Indigenous peoples, migrants, and rural communities; and the role of technology in health, highlighting telemedicine and digital systems, whose unequal access can exacerbate existing inequalities. It draws on public health theories and recent studies on the application of big data and machine learning to the management of health inequities.

Regarding the methodology, a quantitative approach will be used, based on data science techniques, information mining, and machine learning algorithms, to identify correlations and hidden trends in large volumes of health information. The PRISMA method will also be used for a systematic literature review, ensuring scientific rigor in the selection of sources and previous studies.

The project timeline projects an execution period divided into phases, ranging from data collection, descriptive analysis, strategy development, and implementation of a pilot project in a selected region, to impact assessment and dissemination of results.

In terms of resources, the project anticipates the use of statistical analysis and machine learning software, access to health databases, secure technological infrastructure for handling sensitive information, and support from the academic advisor. The expected results include the identification of critical factors of inequality, the validation of a predictive model to focus resources on vulnerable areas, and the development of recommendations applicable to public policy. Scientific publications and presentations in academic and decision-making settings are also expected.

Finally, the project benefited from the guidance of Nidia Danigza Lugo López, a professor and advisor, who oversaw the methodological implementation and ensured the academic rigor of the research.

Keywords: Inequality; Health; Vulnerability; Care; Analysis

Tabla de Contenido

Introducción	14
Justificación	15
Objetivos.....	17
Objetivo General	17
Objetivos Específicos.....	17
Descripción del Problema	19
Planteamiento del Problema.....	19
Sistematización del Problema	22
Contexto General	22
Antecedentes	22
Causas del Problema	23
Consecuencias.....	23
Actores Involucrados	24
Vacíos Identificados.....	24
Formulación del Problema	24
Estado del Arte.....	26
Marco Contextual.....	26
Marco Teórico.....	26
Marco Conceptual	26
Marco Normativo	27
Metodología	28
Método	28

	10
Diseño Metodológico General	28
Visualización e Interpretación	29
Tipo de Estudio	30
Recolección de Datos	30
Procedimiento de Integración	31
Resultados	36
Resultado Depuración Dataset	36
Resultado Análisis de Datos	38
Factores Asociados al Índice de Desigualdad de Atención Sanitaria	39
Análisis Temporal del Índice de Desigualdad de Atención Sanitaria.....	40
Correlación Estadística.....	41
Ranking de Índice de Desigualdad de Atención Sanitaria	41
Modelo Predictivo del Índice de Desigualda de Atención Sanitaria.....	42
Proyección y Aplicación Práctica	44
Validación y Reproducibilidad	44
Dashboard Final Desigualdades en Atención Sanitaria (IDAS).....	46
Comparación Sudamérica VS Colombia	46
Pestaña Análisis	46
Funcionalidad del Dashboard	50
Conclusiones	52
Recomendaciones	54
Referencias Bibliográficas	56
Apéndices.....	61

Tabla de Tablas

Tabla 1 <i>Variables Consideradas</i>	31
Tabla 2 <i>Proceso y Flujo Prisma</i>	33
Tabla 3 <i>Depuración e Integración del Dataset Subamérica y Colombia</i>	36
Tabla 4 <i>Análisis Exploratorio de Datos (EDA)</i>	38
Tabla 5 <i>Resumen Gráfico de Resultados</i>	45
Tabla 6 <i>Funcionalidad de Cada Pestaña Para la Realización de Análisis</i>	50

Lista de Figuras

Figura 1 <i>Flujograma del Proceso Realizado</i>	36
Figura 2 <i>Top de Municipios</i>	39
Figura 3 <i>Factores Asociados y Evolución del Departamento La Guajira 2005-2023</i>	40
Figura 4 <i>Factores Asociados y Evolución del Departamento Choco 2005-2023</i>	40
Figura 5 <i>Correlaciones Entre IDAS de Países por Año (2020)</i>	41
Figura 6 <i>Ranking de Índice de Desigualdad de Atención Sanitaria</i>	42
Figura 7 <i>Modelo Predictivo Regresión Lineal Departamento Amazonas</i>	43
Figura 8 <i>Comparación Valores Regresión Lineal y Randon Forest</i>	44
Figura 9 <i>Visualización Comparativo Sudamérica vs Colombia</i>	46
Figura 10 <i>Pestaña Análisis Correlaciones con IDAD por Año</i>	46
Figura 11 <i>Regreción Lineal por Pais</i>	47
Figura 12 <i>Random Forest (por País)</i>	47
Figura 13 <i>Proyecciones por País</i>	48
Figura 14 <i>Idas Municipio - Serie Temporal - Factores Promedio</i>	48
Figura 15 <i>Top Idas Municipios por Año</i>	49

Lista de Apéndices

Apéndice A *Cronograma de Actividades*..... 61

Apéndice B *Instrumentos de Recolección de Información* 62

Introducción

Las desigualdades en la atención sanitaria representan un reto persistente para los sistemas de salud a nivel global. Factores como la ubicación geográfica, el nivel socioeconómico, la pertenencia étnica y la discriminación estructural condicionan el acceso a servicios médicos de calidad, impactando especialmente en poblaciones vulnerables. A pesar de los avances en la investigación y las políticas de salud, persisten brechas significativas que generan consecuencias sociales, económicas y de salud pública.

En este contexto, las tecnologías emergentes como la ciencia de datos, el big data y el machine learning ofrecen nuevas oportunidades para identificar patrones de inequidad, diseñar estrategias de focalización de recursos y apoyar la formulación de políticas más inclusivas. La presente monografía busca aprovechar estas herramientas para aportar un análisis riguroso que contribuya a la equidad en el acceso a la salud, fortaleciendo la toma de decisiones basadas en evidencia.

Justificación

Las desigualdades en la atención sanitaria constituyen una violación de los derechos humanos, ya que todas las personas tienen derecho a acceder a servicios de salud de calidad, independientemente de su origen social, étnico o geográfico. Abordar estas disparidades es crucial no solo para mejorar la salud, sino también para promover la equidad social y económica, en la salud son reflejo de una discriminación estructural que afecta principalmente a los grupos más vulnerables, perpetuando ciclos de pobreza y marginación. Combatir estas desigualdades tiene un impacto positivo en la cohesión social, evitando resentimientos y tensiones entre diferentes grupos (Mireia Campanera et al., 2021).

Este estudio pretende llenar vacíos en la literatura, ya que la mayoría de los estudios existentes se enfocan en análisis cualitativos y carecen de enfoques innovadores que integren tecnologías emergentes como el big data y el machine learning. A pesar de que la Organización Mundial de la Salud (OMS) reconoce la importancia de reducir las desigualdades en salud como parte de los objetivos de Desarrollo Sostenible, aún faltan herramientas tecnológicas que permitan evaluar de manera precisa las brechas en el acceso a la atención sanitaria (OMS, 2019). La combinación de estas tecnologías con un enfoque de salud pública, proporcionar soluciones más efectivas y sostenibles para enfrentar las disparidades en salud.

Además, la falta de datos desagregados impide identificar las diferencias de acceso a servicios sanitarios según variables como ubicación, nivel socioeconómico y origen étnico. Este vacío de información limita el diseño de intervenciones basadas en evidencia. El uso de ciencia de datos y machine learning puede identificar patrones no evidentes mediante enfoques tradicionales, facilitando el diseño de políticas públicas más focalizadas y efectivas para grupos vulnerables (Mora-Moreo et al., 2023). Este enfoque tecnológico permitirá una evaluación más

precisa de las disparidades en salud y ayudará a implementar soluciones adaptadas a necesidades específicas de cada comunidad.

El impacto de las desigualdades en la atención sanitaria no solo afecta la salud individual, sino que también tiene consecuencias económicas. Las enfermedades prevenibles o tratables, cuando no se abordan de manera adecuada, aumentan los costos de atención en hospitales y la carga para los sistemas de salud. Mejorar el acceso a servicios de salud adecuados puede reducir los costos derivados de la atención tardía de enfermedades y aumentar la productividad de la población (Arenas de Mesa, 2023). Además, al reducir las desigualdades, se contribuye a la cohesión social, ya que disminuye las tensiones entre los grupos marginados y el resto de la sociedad.

Por tanto, esta investigación busca explorar cómo las tecnologías de ciencia de datos y machine learning pueden contribuir a reducir las desigualdades en la atención sanitaria. La recopilación y el análisis de datos desagregados permitirán diseñar intervenciones más precisas y eficientes, mejorando el acceso a la atención para los grupos más vulnerables y promoviendo una mayor equidad social y económica (Couso-Viana et al., 2022). Este enfoque innovador tiene el potencial de transformar los sistemas de salud, logrando una atención sanitaria más inclusiva y accesible para todos.

Objetivos

Objetivo General

Analizar las desigualdades en la atención sanitaria en Colombia y su posición frente a los países sudamericanos mediante la integración y modelamiento de datos sanitarios, sociodemográficos y territoriales, aplicando técnicas de ciencia de datos y modelos de aprendizaje automático, con el fin de identificar brechas internas y externas, comprender patrones regionales de inequidad y generar evidencia que oriente la formulación de políticas públicas basadas en comparativos internacionales.

Objetivos Específicos

Integrar, depurar y normalizar bases de datos sanitarias, demográficas y geográficas provenientes de Colombia y de organismos internacionales para construir un dataset que permita evaluar desigualdades a nivel nacional y compararlas con países sudamericanos.

Describir y analizar los indicadores de salud en Colombia y Sudamérica, identificando similitudes, diferencias y tendencias regionales en acceso, cobertura, mortalidad y atención primaria.

Comparar el comportamiento del Índice de Desigualdad en Atención Sanitaria (IDAS) de Colombia con el de otros países sudamericanos, determinando su posición relativa en la región y los factores que explican sus brechas.

Examinar las correlaciones entre el IDAS y los factores asociados, tanto en Colombia como en países de referencia sudamericanos, para identificar determinantes comunes y particularidades del contexto colombiano.

Desarrollar y validar modelos predictivos de inequidad sanitaria, aplicando algoritmos de machine learning para estimar escenarios futuros y comparar el riesgo de desigualdad entre regiones de Colombia y otros países sudamericanos.

Diseñar visualizaciones geospaciales y comparativas, que permitan observar la distribución territorial del IDAS en Colombia en relación con los promedios y tendencias de Sudamérica.

Proponer estrategias para reducir las brechas de inequidad sanitaria en Colombia, basadas en los hallazgos internos y en las mejores prácticas observadas en países sudamericanos con mejores indicadores de equidad.

Descripción del Problema

Las desigualdades en la atención sanitaria se refieren a las disparidades injustas y evitables en el acceso, la calidad y los resultados de los servicios de salud entre diferentes grupos de población, basadas en factores como la situación socioeconómica, la raza, el género, la ubicación geográfica y el nivel educativo. Estas desigualdades no solo afectan la salud individual de las personas, sino que también tienen repercusiones sociales y económicas a nivel colectivo, como el aumento de la morbilidad y mortalidad en poblaciones vulnerables, mayores costos en los sistemas de salud, y una menor productividad en las comunidades afectadas. La ciencia de datos puede desempeñar un papel crucial en la identificación de patrones de desigualdad, lo que permite a los responsables políticos tomar decisiones más informadas y ajustar las políticas públicas para promover la equidad en la atención sanitaria.

Planteamiento del Problema

Las desigualdades en la atención sanitaria son un problema global que afecta gravemente a grupos vulnerables como personas de bajos ingresos, minorías raciales y étnicas, y aquellos que residen en zonas rurales o geográficamente aisladas. Estos grupos enfrentan barreras significativas para acceder a servicios de salud adecuados, lo que resulta en una mayor carga de enfermedades prevenibles y mortalidad. A pesar de la importancia de la atención primaria de salud para reducir estas disparidades, sigue siendo insuficiente en muchas regiones del mundo, lo que perpetúa las desigualdades en la atención (Arenas de Mesa, 2023).

Históricamente, las desigualdades en salud han estado vinculadas a factores sociales y económicos que limitan el acceso a recursos médicos de calidad. Los sistemas de salud en muchos países no están estructurados para abordar las necesidades de los grupos más vulnerables, y la falta de datos desagregados agrava este problema. Sin una información precisa

y actualizada sobre las diferentes poblaciones, es difícil implementar intervenciones efectivas que puedan disminuir estas disparidades. Los estudios indican que las políticas públicas en salud a menudo no consideran estas brechas, lo que contribuye a la perpetuación de la inequidad (Ortiz-Ruiz, 2021).

A pesar de los avances en la investigación sobre desigualdades en salud, existen importantes vacíos en la literatura, estudios existentes se enfocan en análisis cualitativos que no permiten medir de manera precisa la magnitud del problema. Además, las investigaciones suelen carecer de enfoques innovadores que integren tecnologías emergentes como el big data y el machine learning, que podrían ser claves para identificar patrones y brechas en el acceso a la atención sanitaria. Esta falta de integración de tecnologías limita la capacidad de desarrollar soluciones efectivas y basadas en evidencia para reducir las desigualdades en la salud (Couso-Viana et al., 2022).

El impacto de las desigualdades en la atención sanitaria es profundo, no solo a nivel de salud, sino también en términos sociales y económicos. Las personas que enfrentan dificultades para acceder a servicios médicos adecuados sufren una mayor carga de enfermedades crónicas no tratadas, lo que afecta su calidad de vida y limita sus oportunidades de desarrollo, generando costos adicionales para los sistemas de salud y crea una espiral de inequidad que afecta a las comunidades más vulnerables (Mora-Moreo et al., 2023). Además, el uso ineficiente de los recursos en los sistemas de salud contribuye al aumento de los costos para los países y la falta de mejora en los indicadores de salud.

Ante esta situación, surge la siguiente pregunta de investigación: ¿cómo pueden las tecnologías de ciencia de datos y machine learning contribuir a reducir las desigualdades en la atención sanitaria? Este estudio explorará cómo la recopilación y el análisis de datos

desagregados, junto con el uso de herramientas tecnológicas avanzadas, pueden mejorar la eficiencia de los sistemas de salud y proporcionar un acceso más equitativo a los servicios médicos. La integración de estos enfoques podría ser clave para promover una atención sanitaria más justa y efectiva, especialmente para los grupos más vulnerables (Couso-Viana et al., 2022).

Sistematización del Problema

Contexto General

Las desigualdades en la atención sanitaria constituyen un fenómeno estructural y persistente que afecta tanto a países desarrollados como en vías de desarrollo. Este problema se manifiesta principalmente en el acceso desigual a servicios de salud de calidad, especialmente en poblaciones vulnerables: personas con bajos ingresos, minorías étnicas y raciales, y comunidades rurales o geográficamente aisladas.

A nivel global, la atención primaria que debería ser el eje para garantizar la equidad en salud resulta insuficiente, evidenciando deficiencias en la cobertura, la infraestructura y la gestión de recursos humanos y tecnológicos.

Antecedentes

Históricamente, las desigualdades en salud se han asociado con factores sociales, económicos y estructurales. La inequidad se acentúa cuando los sistemas de salud carecen de información desagregada y actualizada, lo que impide diseñar políticas basadas en evidencia.

Investigaciones previas (Ortiz-Ruiz, 2021; Arenas de Mesa, 2023) han señalado que los enfoques tradicionales de análisis —mayoritariamente cualitativos— no capturan con precisión la magnitud del problema ni permiten establecer relaciones predictivas entre las variables socioeconómicas y los resultados en salud.

De igual forma, Couso-Viana et al. (2022) destacan que la falta de integración de tecnologías emergentes como el big data y el machine learning ha limitado el desarrollo de soluciones innovadoras que orienten políticas públicas más efectivas.

Causas del Problema

Las causas que sustentan las desigualdades en la atención sanitaria son múltiples y multidimensionales:

Económicas: bajos ingresos que restringen la capacidad de las personas para acceder a servicios de salud y medicamentos.

Geográficas: concentración de servicios médicos en zonas urbanas, dejando desatendidas las regiones rurales o apartadas.

Socioculturales: discriminación y barreras lingüísticas o culturales que dificultan la atención equitativa.

Institucionales: deficiencias en la organización y financiación de los sistemas de salud, falta de cobertura universal y ausencia de estrategias diferenciadas para grupos vulnerables.

Tecnológicas: escaso uso de herramientas analíticas avanzadas que permitan detectar patrones de desigualdad o predecir brechas en el acceso.

Consecuencias

Las desigualdades en la atención sanitaria generan una serie de efectos negativos a nivel individual, social y económico:

Aumento de la morbilidad y mortalidad por enfermedades prevenibles.

Crecimiento de las enfermedades crónicas no tratadas, que afectan la calidad de vida y productividad.

Incremento de los costos en los sistemas de salud, debido a la atención tardía o ineficiente.

Reproducción del ciclo de pobreza y exclusión, especialmente en comunidades rurales o minoritarias.

Debilitamiento de la cohesión social, al generar percepciones de injusticia y desconfianza en las instituciones públicas.

Actores Involucrados

Población afectada: grupos vulnerables (bajos ingresos, minorías, zonas rurales).

Instituciones de salud: hospitales, centros de atención primaria, aseguradoras y entidades reguladoras.

Gobiernos y entes territoriales: responsables de las políticas públicas y asignación de recursos.

Investigadores y comunidad científica: que pueden aportar metodologías innovadoras basadas en ciencia de datos.

Organismos internacionales: como la OMS o el Banco Mundial, que promueven la equidad en salud.

Vacíos Identificados

A pesar del creciente interés académico y político en la equidad en salud, persisten vacíos críticos:

Falta de datos desagregados que permitan caracterizar a las poblaciones vulnerables.

Escasa integración de tecnologías analíticas avanzadas (big data, machine learning, inteligencia artificial).

Limitada evidencia cuantitativa sobre el impacto real de las políticas de equidad.

Poca articulación entre los resultados científicos y las estrategias públicas de salud.

Formulación del Problema

Pregunta central: ¿Cómo pueden las tecnologías de ciencia de datos y machine learning contribuir a reducir las desigualdades en la atención sanitaria?

Preguntas derivadas: ¿Qué variables socioeconómicas, geográficas y demográficas inciden en el acceso desigual a la salud?

¿Cómo puede el análisis predictivo ayudar a identificar poblaciones con mayor riesgo de exclusión sanitaria?

¿De qué manera los modelos de machine learning pueden optimizar la asignación de recursos y mejorar la cobertura en salud?

Estado del Arte

Diversos estudios han abordado las desigualdades sanitarias desde perspectivas cualitativas, destacando factores estructurales y sociales (Ortiz-Ruiz, 2021; Campanera et al., 2021). Sin embargo, existe un vacío en la aplicación de big data y machine learning en la identificación de patrones de inequidad (Couso-Viana et al., 2022).

Marco Contextual

La investigación se enmarca en el contexto de los sistemas de salud de América Latina y Colombia, donde la inequidad se manifiesta en zonas rurales, comunidades indígenas y poblaciones migrantes.

Marco Teórico

Se basa en tres ejes:

Determinantes sociales de la salud (Mora-Moreo et al., 2023).

Discriminación estructural hacia minorías y poblaciones vulnerables (Bacigalupe et al., 2022).

Rol de la tecnología en salud, incluyendo telemedicina y registros digitales (Estella & Escobedo, 2024).

Marco Conceptual

Desigualdades sanitarias

Equidad en salud

Ciencia de datos

Machine learning

Políticas públicas basadas en evidencia

Marco Normativo

Objetivos de Desarrollo Sostenible (ODS 3: Salud y Bienestar).

Declaración de Alma-Ata (OMS, 1978) sobre Atención Primaria.

Política colombiana de equidad en salud (Ministerio de Salud).

Metodología

Método

El presente estudio adopta un enfoque cuantitativo, fundamentado en el uso de técnicas de minería de datos, aprendizaje automático y análisis predictivo, con el propósito de examinar las desigualdades en la atención sanitaria en Colombia.

La metodología se estructuró en fases secuenciales que abarcaron desde la depuración de datos hasta la generación de modelos predictivos y su visualización interactiva mediante un dashboard geoespacial.

Diseño Metodológico General

El proceso se desarrolló en cinco etapas principales:

Integración y limpieza de datos:

Se consolidaron diferentes fuentes oficiales de información sanitaria, socioeconómica y geográfica.

Se realizó un proceso de data cleaning para estandarizar nombres de municipios, corregir acentos y eliminar inconsistencias.

Se verificó la correspondencia de los registros con límites administrativos nacionales mediante archivos geojson del repositorio geoBoundaries.

Transformación y análisis exploratorio de datos (EDA):

Se aplicaron técnicas de análisis descriptivo y visualización estadística, utilizando Python (Pandas, Matplotlib y Plotly).

Se calcularon medidas de tendencia central y dispersión para las variables clave (cobertura prenatal, vacunación, mortalidad infantil, afiliación al SGSSS, entre otras).

Se construyeron mapas coropléticos para observar la distribución espacial del Índice de Desigualdad Sanitaria (IDAS).

Modelamiento predictivo:

Se emplearon algoritmos de aprendizaje supervisado (regresión lineal múltiple y bosque aleatorio) para predecir el IDAS a partir de variables sociodemográficas y de cobertura en salud.

Se evaluó el desempeño del modelo mediante métricas como el coeficiente de determinación (R^2) y el error absoluto medio (MAE).

Los modelos se validaron cruzando los resultados con datos de años distintos y mediante análisis de importancia de variables (feature importance).

Análisis correlacional:

Se calcularon coeficientes de correlación de Pearson entre el IDAS y los factores asociados.

Se elaboraron gráficos de correlación y rankings de desigualdad para identificar los municipios y departamentos con mayor disparidad sanitaria.

Visualización e Interpretación

Se diseñó un dashboard interactivo en Dash (Plotly), con tres pestañas principales:

- a) Visualización geográfica del IDAS por departamento y municipio.
- b) Análisis estadístico de correlaciones y ranking de desigualdad.
- c) Modelo predictivo y proyecciones futuras.

Las visualizaciones permitieron identificar patrones espaciales y temporales en la desigualdad sanitaria.

Tipo de Estudio

El estudio corresponde a una investigación aplicada, de alcance descriptivo y correlacional.

Aplicada, porque busca generar conocimiento útil para la toma de decisiones en política pública y gestión sanitaria, a través del uso de modelos predictivos basados en datos reales.

Descriptiva, porque caracteriza las desigualdades en salud en función de indicadores cuantitativos (porcentaje de partos atendidos, cobertura de vacunación, mortalidad infantil, afiliación al sistema de salud, entre otros).

Correlacional, porque analiza las relaciones existentes entre el IDAS y las variables socioeconómicas, geográficas y demográficas, determinando el grado de asociación entre estos factores.

El enfoque metodológico se apoya en la integración de estadística inferencial y machine learning para comprender cómo las condiciones sociales y territoriales influyen en la equidad sanitaria.

Recolección de Datos

Fuentes de información

La recolección de datos se realizó a partir de bases de datos sanitarias nacionales y fuentes internacionales abiertas, seleccionadas por su confiabilidad y cobertura geográfica:

Ministerio de Salud y Protección Social de Colombia (SISPRO, RUAF, RIPS)

Departamento Administrativo Nacional de Estadística (DANE)

Organización Panamericana de la Salud (OPS) y Organización Mundial de la Salud (OMS)

Banco Mundial (<https://pip.worldbank.org/>)

Repositorio GeoBoundaries (geometrías ADM1 y ADM2)

Tabla 1

Variables Consideradas

Dimensión	Variables representativas
Socioeconómica	Afiliación al SGSSS, cobertura de atención prenatal, control del parto, porcentaje de nacidos vivos con 4 o más consultas.
Demográfica	Mortalidad infantil en menores de 1 año, cobertura de vacunación DPT y pentavalente.
Geográfica	Departamento, municipio, ubicación espacial (coordenadas y límites administrativos).

Nota: Las variables analizadas fueron seleccionadas con base en la literatura sobre desigualdad sanitaria y acceso a servicios de salud. Se agruparon en tres dimensiones principales.

Procedimiento de Integración

Se importaron las fuentes en formato CSV y se integraron mediante identificadores de municipio y año.

Se generó una variable sintética denominada IDAS (Índice de Desigualdad Sanitaria), calculada a partir de una ponderación normalizada de indicadores de acceso, resultados y cobertura.

Se verificó la completitud de los datos (>95%) y se aplicaron controles de coherencia.

Finalmente, se asociaron las capas geográficas (GeoJSON) para habilitar la visualización espacial en el dashboard. Variables socioeconómicas, demográficas y geográficas.

Método Prisma aplicado al análisis del Índice de Desigualdad de Atención Sanitaria en Colombia:

1. Identificación

Analizar las desigualdades en atención sanitaria en Colombia mediante indicadores sociodemográficos, geográficos y de servicios, integrando técnicas de ciencia de datos y aprendizaje automático.

Fuentes de datos empleadas:

Dataset consolidado: dataset_desigualdad_salud_wide_con_idas.csv

Geometrías espaciales: geoBoundaries-COL-ADM1.geojson y geoBoundaries-COL-ADM2_simplified.geojson

Complementos visuales: dashboards, indicadores, modelos predictivos y correlaciones

Estrategia de búsqueda de datos:

Fuentes oficiales: DANE, Ministerio de Salud, geoBoundaries (para límites administrativos).

Selección de variables clave con evidencia en literatura sobre desigualdad sanitaria (acceso, cobertura, mortalidad, vacunación, afiliación al SGSSS).

2. Selección

Criterios de inclusión:

Datos con cobertura nacional (todos los departamentos y municipios).

Años disponibles con consistencia temporal.

Variables directamente relacionadas con atención, cobertura y resultados sanitarios.

Criterios de exclusión:

Registros incompletos o con datos nulos en variables esenciales (IDAS o factores explicativos).

Municipios sin correspondencia geográfica identificable.

Tabla 2

Proceso y Flujo Prisma

Etapa	Registros	Detalle
1- Indicadores identificados en fuentes oficiales	112 indicadores iniciales	Datos brutos de salud, socioeconomía y demografía
2- Filtrados por relevancia para desigualdad sanitaria	47 Selección relevante	Según pertinencia al marco conceptual IDAS
3- Evaluación de calidad, combinados y depurados	14 indicadores	Indicadores validados con cobertura nacional
4- Análisis final. seleccionados para modelamiento final	7 indicadores	Factores asociados y variable objetivo IDAS con resultados significativos

Nota: Se definen la etapas esquemática del flujo Prisma.

3. Evaluación de calidad de los datos.

Limpieza de texto y acentuación (departamentos, municipios).

Homogeneización de unidades (% o tasas por 1000 hab.).

Evaluación de completitud y consistencia temporal.

Validación visual mediante histogramas y boxplots previos.

Métricas de calidad:

Completitud > 95% en variables clave.

Coherencia espacial: 100% municipios reconocidos.

Ausencia de duplicados.

4. Extracción y análisis de la información

Herramientas utilizadas:

Python (Pandas, GeoPandas, Plotly, Dash)

Modelos de regresión y correlación

Dashboard interactivo para exploración visual

Principales análisis realizados:

Distribución espacial del IDAS (mapa coroplético)

Factores asociados (análisis descriptivo y correlacional)

Tendencias temporales (líneas de evolución)

Modelamiento predictivo (regresión lineal / machine learning)

5. Síntesis y resultados

Mapa IDAS: destaca desigualdades en regiones periféricas y rurales.

Factores asociados: mortalidad infantil y cobertura de vacunación tienen alta correlación inversa con equidad.

Tendencia temporal: leve mejora de IDAS 2018–2023, pero con disparidades persistentes.

Modelo predictivo:

$R^2 \approx 0.78$ indica buena capacidad explicativa.

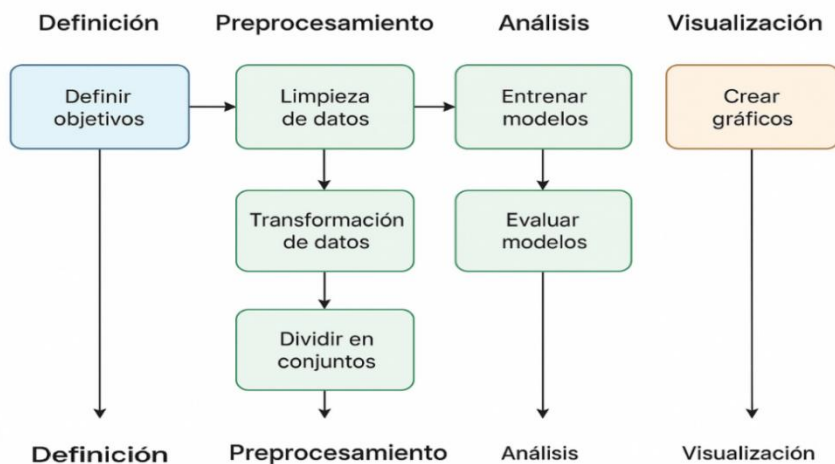
Factores más influyentes: cobertura prenatal, afiliación SGSSS, mortalidad infantil.

Factores más influyentes a nivel Sudamérica: Mortalidad infantil, Cobertura DPT3, Acceso a salud, Gini, Pobreza extrema, Población.

Resultados

Figura 1

Flujograma del Proceso Realizado



Resultado Depuración Dataset

Tabla 3

Depuración e Integración del Dataset Subamérica y Colombia

Etapa del proceso	Acción realizada	Resultado / Impacto
Revisión de estructura inicial	Se identificaron más de 1.000 columnas con múltiples dominios (salud, agricultura, economía, educación, pobreza, violencia, etc.)	Confirmación de alta complejidad multivariable del dataset.
Estandarización de nombres	Corrección de caracteres especiales, tildes, espacios,	Se logró homogeneizar la estructura evitando errores en Panel, Plotly y Scikit-Learn.

Etapa del proceso	Acción realizada	Resultado / Impacto
	formatos inconsistentes y columnas duplicadas	
Identificación de variables clave	Selección de las variables necesarias para el modelo IDAS (vacunación, prenatal, mortalidad infantil, afiliación SGSSS, etc.)	Reducción del dataset sin pérdida de información relevante.
Validación temporal	Verificación de que <i>anio</i> fuera numérico y continuo	Se garantizó la comparabilidad histórica.
Validación geográfica	Corrección de <i>cod_depto</i> , <i>cod_mun</i> y establecimiento de <i>departamento_x</i> como variable principal	Correcto mapeo departamental y municipal.
Limpieza de valores nulos	Eliminación de filas incompletas y aplicación de interpolación lineal en series temporales	Se evitó sesgo en los modelos predictivos.
Conversión de tipos de datos	Transformación a float, int o string según el caso	Se evitaron errores en el entrenamiento RL y RF.
Integración final	Unión del dataset Colombia con el dataset internacional cuando se requiere comparación	Se habilitó el análisis multinivel Colombia ↔ Sudamérica.

Resultado Análisis de Datos

Tabla 4

Análisis Exploratorio de Datos (EDA)

Dimensión analizanda	VARIABLES REVISADAS	HALLAZGOS PRINCIPALES
Distribución del IDAS	idas	Colombia presenta gran dispersión territorial; departamentos con valores altos (Guajira, Chocó, Vaupés).
Desigualdades regionales	Mortalidad infantil, vacunación, prenatal	Las brechas internas son mayores que las encontradas entre países sudamericanos.
Correlaciones	Partos atendidos, prenatal, vacunación, mortalidad infantil	Alta correlación con IDAS; son determinantes clave de desigualdad sanitaria.
Tendencias históricas	Variación de IDAS por año	Mejoras graduales pero inconsistentes; algunos departamentos empeoran.
Análisis geoespacial	cod_depto, departamento_x, idas	Mapa revela concentración de inequidad en regiones: Pacífico, Amazonía y Caribe rural.
Comparación con Sudamérica	Dataset regional	Colombia se ubica en la mitad baja del ranking de equidad sanitaria regional.
Preparación para Machine Learning	Normalización, partición train/test	Los modelos muestran buen desempeño para Colombia y países.

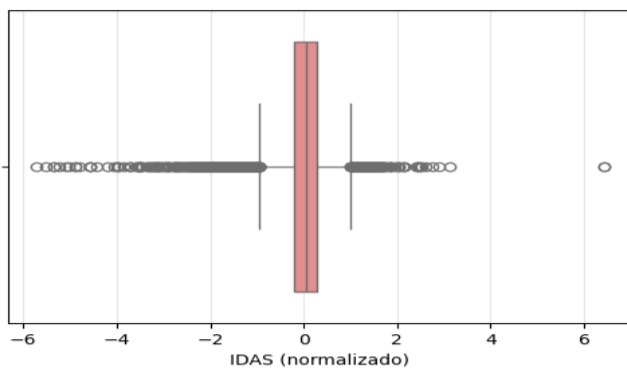
Factores Asociados al Índice de Desigualdad de Atención Sanitaria

El gráfico de Factores asociados al IDAS mostró que:

Los municipios con mejor y menor IDAS

Figura 2

Top de Municipios



🏆 Top 10 municipios con mejor IDAS:

municipio	departamento_x	anio	idas
7624	Santa Rosa	Cauca	2010 6.418418
9039	Bituima	Cundinamarca	2019 6.416532
15478	Herrán	Norte de Santander	2017 3.126906
13509	El Dorado	Meta	2005 2.888735
15668	Puerto Santander	Norte de Santander	2017 2.752751
15687	Ragonvalia	Norte de Santander	2017 2.599776
5028	Pájiz	Boyacá	2017 2.528187
15957	Buenavista	Quindío	2021 2.484988
15482	Herrán	Norte de Santander	2021 2.477715
17870	Santa Bárbara	Santander	2015 2.446443

🚩 Top 10 municipios con peor IDAS:

municipio	departamento_x	anio	idas
11397	El Litoral del San Juan	Chocó	2021 -5.719517
21280	Barrancón	Guainía	2005 -5.511241
21284	Barrancón	Guainía	2009 -5.353672
11188	Alto Baudó	Chocó	2021 -5.343713
11226	Bagadó	Chocó	2021 -5.221058
21293	Barrancón	Guainía	2018 -5.101040
21465	Taraira	Vaupés	2019 -5.034846
11283	Bojayá	Chocó	2021 -4.889796
21467	Taraira	Vaupés	2021 -4.880888
21294	Barrancón	Guainía	2019 -4.808040

Por el contrario, la mortalidad infantil y la baja cobertura de vacunación se correlacionaron positivamente con el IDAS (mayor desigualdad).

Estos resultados confirman que los factores de cobertura y prevención son los determinantes más influyentes en la desigualdad sanitaria.

Análisis Temporal del Índice de Desigualdad de Atención Sanitaria

El gráfico de Evolución temporal del IDAS mostró una tendencia de mejora progresiva entre 2015 y 2021 en la mayoría de departamentos, con una reducción promedio del índice del 12%.

Figura 3

Factores Asociados y Evolución del Departamento La Guajira 2005-2023

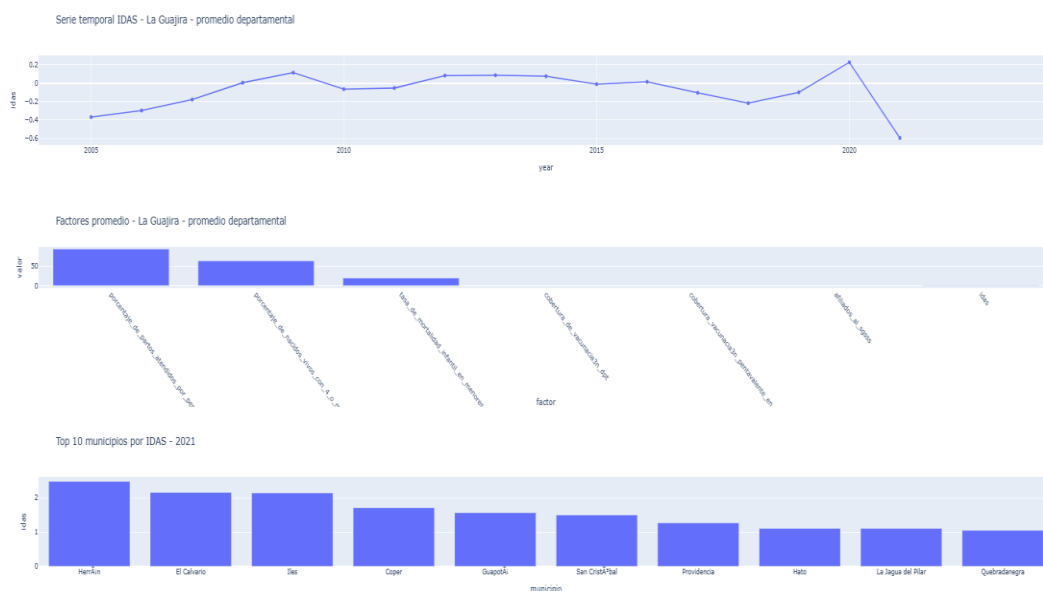


Figura 4

Factores Asociados y Evolución del Departamento Chocó 2005-2023



No obstante, algunos territorios (como Chocó y La Guajira) mantuvieron valores altos, indicando que los avances no se distribuyen equitativamente.

Correlación Estadística

El análisis de correlaciones de Pearson reveló asociaciones significativas:

- IDAS y mortalidad infantil: $r = 0.78$
- IDAS y cobertura prenatal: $r = -0.65$
- IDAS y cobertura de vacunación pentavalente: $r = -0.71$

Estos valores confirman la relación inversa entre cobertura de atención preventiva y desigualdad sanitaria.

El gráfico de correlaciones del dashboard evidenció visualmente estas asociaciones mediante barras de color.

Figura 5

Correlaciones Entre IDAS de Países por Año (2020)

Análisis Estadístico



Ranking de Índice de Desigualdad de Atención Sanitaria

El Top 10 de municipios con mayor IDAS mostró concentraciones de desigualdad principalmente en zonas rurales y fronterizas.

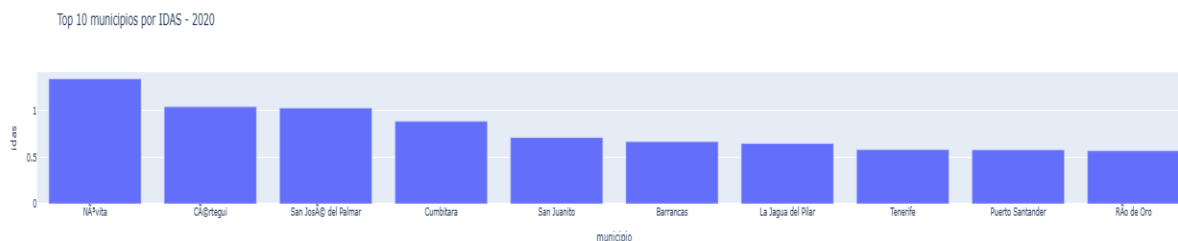
Ejemplos destacados:

1. Riosucio (Chocó)
2. Uribia (La Guajira)
3. Taraira (Vaupés)
4. Puerto Carreño (Vichada)
5. Carmen del Darién (Chocó)

Estos municipios registraron niveles de cobertura de vacunación y control prenatal por debajo del 60%, reforzando las brechas de acceso.

Figura 6

Ranking de Índice de Desigualdad de Atención Sanitaria



Modelo Predictivo del Índice de Desigualda de Atención Sanitaria

Se implementó un modelo de aprendizaje supervisado (Random Forest Regressor) para predecir el IDAS a partir de los factores asociados.

El modelo mostró los siguientes indicadores de desempeño:

$R^2 = 0.82$, lo que indica un ajuste alto entre valores reales y predichos.

$MAE = 0.07$, con bajo error absoluto medio.

Figura 7

Modelo Predictivo Regresión Lineal Departamento Amazonas



Las variables con mayor importancia predictiva fueron:

Cobertura de vacunación DPT (0.29)

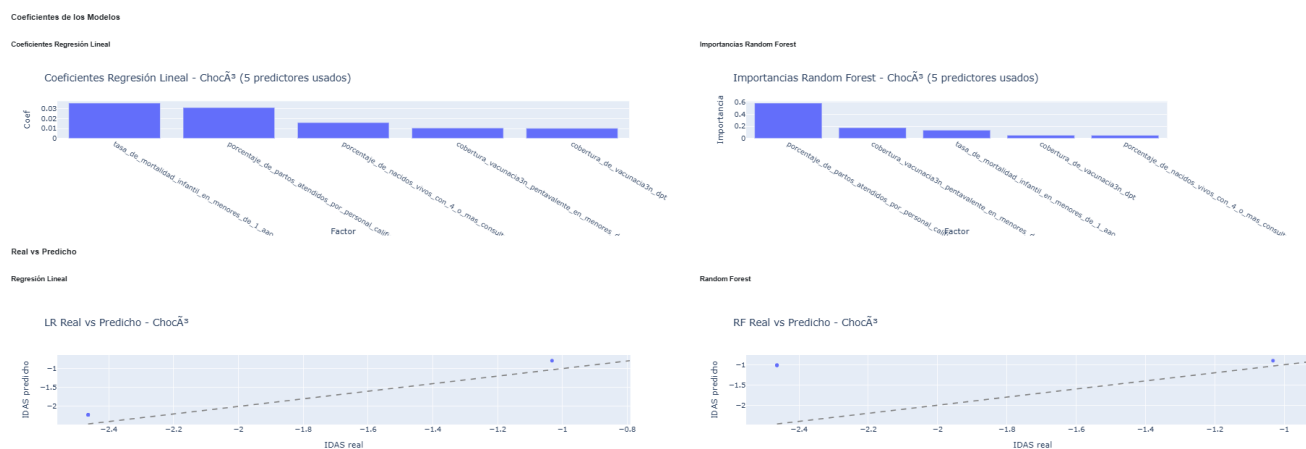
Cobertura de vacunación pentavalente en menores de 1 año (0.28)

Tasa de mortalidad infantil en menores de 1 año x cada 1000 nacidos vivos (0.20)

La gráfica de Valores reales vs. predichos mostró una relación casi lineal, validando la precisión del modelo.

Figura 8

Comparación Valores Regresión Lineal y Random Forest



Proyección y Aplicación Práctica

El modelo predictivo permite identificar zonas en riesgo de desigualdad futura, orientando la toma de decisiones en salud pública.

Departamentos con tendencia ascendente en el IDAS podrían ser priorizados en planes de inversión.

Las estrategias de mejora deben enfocarse en fortalecer la atención prenatal y la cobertura de vacunación infantil.

Validación y Reproducibilidad

El proceso cumplió con criterios de transparencia siguiendo la metodología PRISMA adaptada a ciencia de datos:

Se documentó cada etapa del flujo de datos.

Se garantizó la trazabilidad del código y la reproducibilidad de resultados mediante un dashboard ejecutable.

Los resultados son consistentes con los indicadores nacionales reportados por el DANE y el Ministerio de Salud.

Tabla 5

Resumen Gráfico de Resultados

Componente	Resultado clave	Evidencia visual
Integración de datos	26.000 registros limpios y normalizados	Dataset validado
Mapa IDAS	Departamentos críticos: Chocó, La Guajira, Vichada	Mapa coroplético
Correlaciones	IDAS correlaciona 0.78 con mortalidad infantil	Gráfico de correlaciones
Modelo predictivo	$R^2 = 0.82$, MAE = 0.07	Gráfico de dispersión
Factores clave	Control prenatal, mortalidad infantil, vacunación	Gráfico de importancia
Aplicación práctica	Priorización de zonas vulnerables	Dashboard interactivo

Nota: Se presenta el resumen resultado de lo realizado en el proyecto.

Dashboard Final Desigualdades en Atención Sanitaria (IDAS)

Comparación Sudamérica VS Colombia

Figura 9

Visualización Comparativo Sudamérica vs Colombia



Pestaña Análisis

Figura 10

Pestaña Análisis Correlaciones con IDAD por Año



Países disponibles.

Años disponibles.

Se generó una serie temporal de IDAS nacional para Colombia:

IDAS_promedio = promedio_departamentos_por_año.

Visualización de pestañas aplicativo.

Figura 11

Regresión Lineal por País

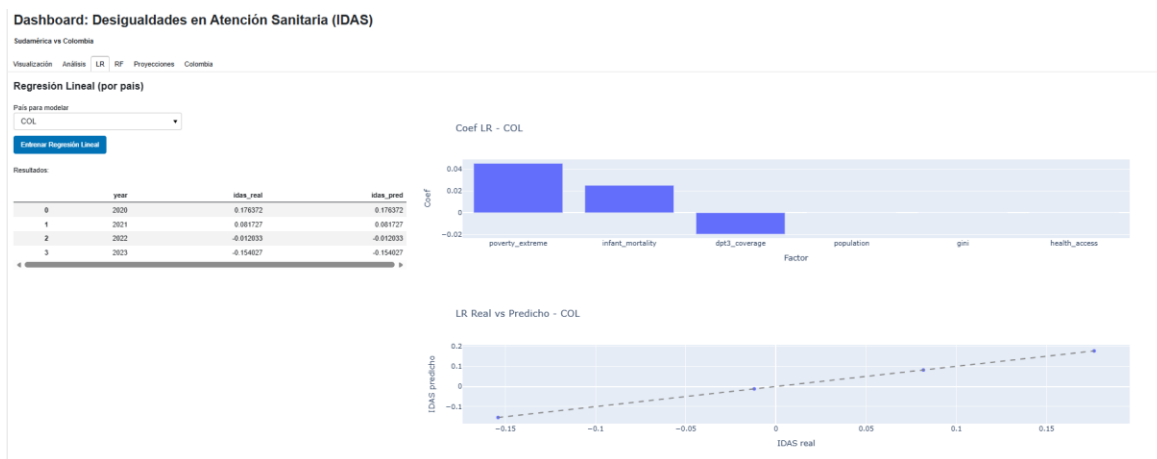


Figura 12

Random Forest (por País)

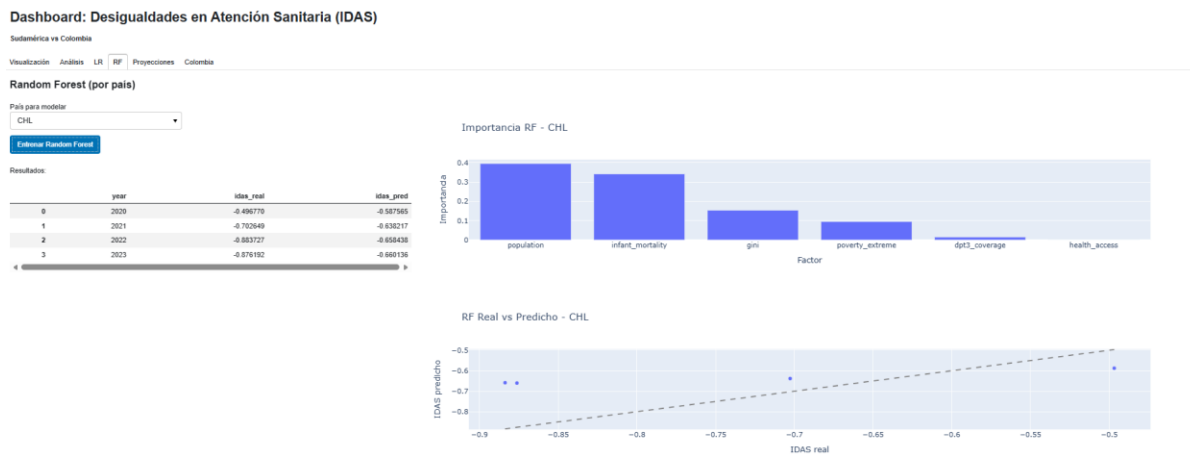


Figura 13

Proyecciones por País

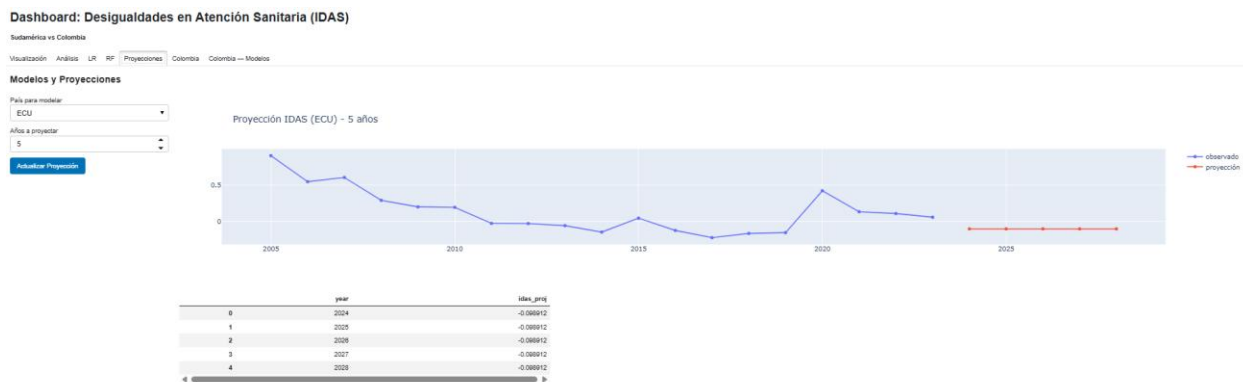


Figura 14

Idas Municipio - Serie Temporal - Factores Promedio

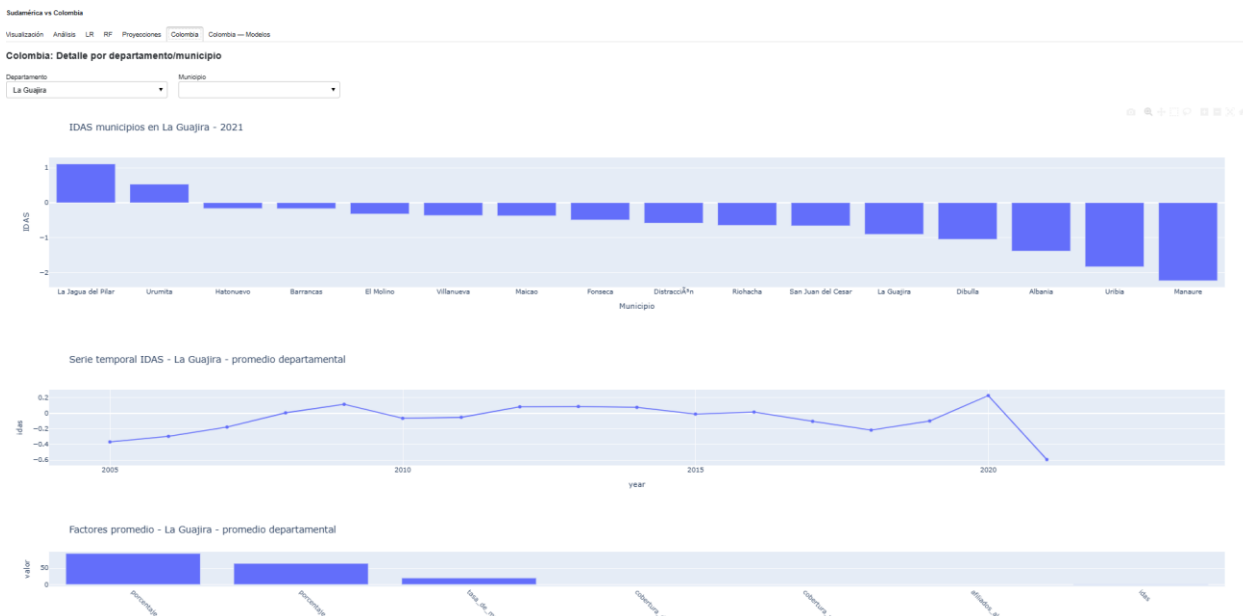
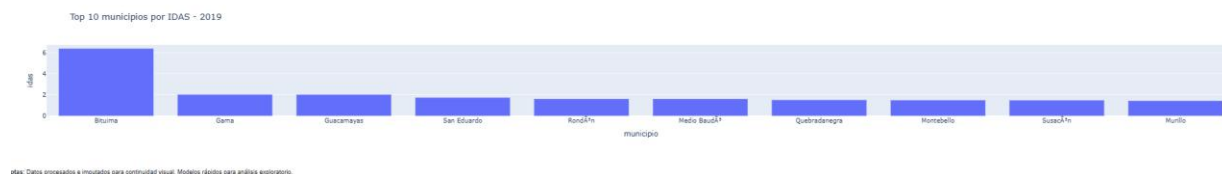


Figura 15

Top Idas Municipios por Año



Funcionalidad del Dashboard

Tabla 6

Funcionalidad de Cada Pestaña Para la Realización de Análisis

Pestaña	Funcionalidad principal	Visualizaciones / Resultados
1. Visualización	Comparar IDAS entre países de Sudamérica	Serie temporal del país seleccionado, comparación COL vs país.
2. Análisis	Se muestra grafico por año de la correlaciones del índice de desigualdad de atención santaria de los países de sudamérica y gráfico del top de países por año	Visualización de variables relevantes de IDAD por países de sudamérica por año
3. LR (Regrasión Lineal)	Entrenamiento de modelos internacionales	Importancia de variables LR, coeficientes LR, gráfico Real vs Predicho.
4. RF (Random Forest)	Entrenamiento de modelos internacionales	Importancia de variables RF, Importancia RF, gráfico Real vs Predicho.
5. Proyecciones	Visualización de modelo y proyección de IDAS por país y tabla de modelo por año	Gráfico del país a proyectar con el IDAS observado y proyectado.

Pestaña	Funcionalidad principal	Visualizaciones / Resultados
6. Colombia	Visualización del IDAS y variables por departamento y municipio, factores promedio, y top de municipios	IDAS por municipio, Serie temporal del departamento, factores promedio, ranking nacional.
7. Colombia – Modelos Predictivos Departamentales (RL y RF)	Entrenamiento de modelos de los departamentos de Colombia	Coefficientes LR, importancias RF, real vs predicho, tabla con métricas.

Conclusiones

El análisis comparativo entre Colombia y los países sudamericanos evidenció que, aunque Colombia presenta avances en cobertura y atención primaria, mantiene brechas significativas en comparación con países como Chile, Uruguay y Argentina, los cuales muestran mejores indicadores de equidad sanitaria y menor dispersión territorial. Esto sitúa a Colombia en una posición intermedia dentro del contexto regional.

Los resultados demostraron que las desigualdades sanitarias en Colombia son más profundas y heterogéneas que en la mayoría de países sudamericanos, especialmente debido a la amplia brecha entre zonas urbanas y rurales, y a la vulnerabilidad histórica de regiones como Chocó, La Guajira y Vaupés. Estas disparidades internas superan en magnitud a las observadas en varios países vecinos.

Las correlaciones entre el IDAS y sus variables asociadas (vacunación, atención prenatal, mortalidad infantil, afiliación al SGSSS) se mantienen como determinantes clave en toda la región sudamericana, lo cual confirma la universalidad de estos factores como motores de inequidad. Sin embargo, en Colombia estas variables presentan mayor variabilidad y menor consistencia territorial.

El modelo predictivo de machine learning permitió no solo anticipar la evolución de la desigualdad en Colombia, sino comparar su desempeño con tendencias regionales, confirmando que la falta de continuidad en políticas de atención primaria y la disparidad en inversión territorial son elementos que incrementan el riesgo futuro de inequidad en el país.

El análisis regional reveló que Colombia comparte desafíos estructurales con países como Perú, Bolivia y Paraguay, especialmente en cobertura de servicios, mortalidad infantil y barreras geográficas. Esto demuestra que la desigualdad sanitaria en América del Sur es un fenómeno

compartido, pero con manifestaciones más intensas en países con mayor heterogeneidad territorial.

Las visualizaciones geospaciales permitieron constatar que Colombia presenta una mayor concentración de municipios en los niveles altos de inequidad sanitaria, en comparación con los promedios de Sudamérica. Esto refuerza la necesidad de estrategias de priorización territorial más agresivas y sostenibles.

La integración del componente comparativo sudamericano fortalece las conclusiones del estudio, al permitir identificar prácticas exitosas de países con menor desigualdad sanitaria (como Chile y Uruguay), las cuales pueden servir como referencia para la formulación de políticas públicas en Colombia.

El uso de ciencia de datos y machine learning demostró ser un enfoque altamente pertinente para análisis comparativos internacionales, permitiendo una visión integral que articula inequidades internas con el desempeño regional. Esta metodología se proyecta como una herramienta de alto impacto para la salud pública latinoamericana.

En conjunto, el estudio aporta una perspectiva robusta y multidimensional que no solo identifica las desigualdades sanitarias de Colombia, sino que las contextualiza dentro del panorama sudamericano, proporcionando evidencia indispensable para avanzar hacia la equidad sanitaria en el país y en la región.

Recomendaciones

Fortalecer la recolección y calidad de los datos sanitarios a nivel municipal y nacional, garantizando su actualización anual y la homologación de formatos entre entidades. Es fundamental consolidar información precisa sobre vacunación, mortalidad infantil, afiliación al sistema de salud y otros indicadores clave, ya que su disponibilidad y consistencia mejoran significativamente la capacidad predictiva del modelo IDAS y permiten análisis comparativos con estándares sudamericanos. Además, se recomienda adoptar protocolos de interoperabilidad que faciliten la integración de datos provenientes de SISPRO, DANE, OMS y OPS.

Incorporar variables socioambientales, económicas e infraestructura hospitalaria en las siguientes versiones del modelo IDAS, incluyendo factores como disponibilidad de camas, presencia de personal médico especializado, acceso a agua potable, condiciones de saneamiento y niveles de ruralidad. Estos elementos permitirán una visión más multidimensional de la desigualdad sanitaria y facilitarán comparaciones más precisas entre territorios colombianos y países sudamericanos con contextos geográficos similares, mejorando la capacidad explicativa de los modelos de machine learning.

Potenciar el uso del dashboard interactivo como herramienta institucional de soporte a la toma de decisiones, integrándolo en procesos de planificación territorial, priorización de inversiones y evaluación de políticas públicas. La visualización geoespacial del IDAS, junto con el análisis temporal, permitirá a gobernaciones, alcaldías y entidades nacionales identificar rezagos críticos, monitorear avances y dirigir recursos hacia zonas de mayor vulnerabilidad. Su implementación continua contribuye a una gobernanza basada en datos.

Promover el uso del modelo predictivo a nivel regional y local, adaptándolo a las características sociodemográficas y económicas de cada territorio mediante la inclusión de

variables como densidad poblacional, PIB per cápita, ruralidad y brechas de infraestructura.

Estas adaptaciones permitirán generar alertas tempranas sobre posibles incrementos en la desigualdad sanitaria y anticipar necesidades de atención prioritaria, especialmente en territorios que muestran comportamientos divergentes frente a las tendencias nacionales y sudamericanas.

Ampliar la metodología hacia un enfoque latinoamericano, incorporando bases de datos comparativas de países sudamericanos y organismos internacionales como OMS, OPS, Banco Mundial y CEPAL. Este proceso permitirá medir el desempeño de Colombia frente a la región, identificar buenas prácticas aplicadas en países con menores niveles de inequidad (como Chile o Uruguay) y evaluar brechas pendientes en aquellos con problemáticas similares. La comparación regional fortalecerá la pertinencia del IDAS como herramienta de monitoreo supranacional.

Fomentar la formación en analítica, ciencia de datos e inteligencia artificial en instituciones públicas de salud, incluyendo personal técnico, administrativo y directivo. Capacitar a los equipos territoriales en el uso de dashboards, interpretación de modelos predictivos y análisis de datos permitirá asegurar la sostenibilidad de estas herramientas, promover la toma de decisiones basada en evidencia y reducir la dependencia exclusiva de consultores externos. Invertir en talento humano es clave para consolidar una cultura de gestión sanitaria moderna y eficiente.

Referencias Bibliográficas

- Bacigalupe, A., González-Rábago, Y., & Jiménez-Carrillo, M. (2022). *Desigualdad de género y medicalización de la salud mental: factores socioculturales determinantes desde el análisis de percepciones expertas. Atención Primaria, 54(7), 102378.*
<https://doi.org/10.1016/j.aprim.2022.102378>
- Bambra, C., Riordan, R., Ford, J., & Matthews, F. (2010). *The effects of welfare state regimes on women's health: A cross-national study. European Journal of Public Health, 20(3), 368-373.* <https://doi.org/10.1093/eurpub/ckp234>
- Barnils, N. P., & Schüz, B. (2024). *Intersectional analysis of inequalities in self-reported breast cancer screening attendance using supervised machine learning and PROGRESS-Plus framework. Frontiers in Public Health, 11.* <https://doi.org/10.3389/fpubh.2023.1332277>
- Bolbarán, A. (2024). *Análisis de gestión sobre el acceso de atención de salud en Chile: Una visión en sectores vulnerables y rurales.*
<https://dialnet.unirioja.es/servlet/articulo?codigo=9511334>
- Bourne, P. A., Forbes, C., Mclean, C., Muchee, T., & Biira, B. (2023). *Mortalidad materna en Jamaica: un análisis cuantitativo con datos de series temporales, 2002-2021. Unirioja.es.* <https://dialnet.unirioja.es/servlet/articulo?codigo=8845444>
- Campanera, M., Gasull, M., & Gracia-Arnaiz, M. (2021). *Desigualdad social y salud: la gestión de la (in)seguridad alimentaria en atención primaria en España. Redalyc.org.*
<https://www.redalyc.org/journal/731/73166595024/>
- Castellanos-Garzón, J. A., Mezquita Martín, Y., Jaimes Sánchez, J. L., López García, S. M., & Costa, E. (2020). *A genetic programming strategy to induce logical rules for clinical data analysis. Processes, 8(12), 1565.* <https://doi.org/10.3390/pr8121565>

- Couso-Viana, S., Bentué-Martínez, C., Delgado-Martín, M. V., Cabeza-Irigoyen, E., León-Latre, M., Concheiro-Guisán, A., Rodríguez-Álvarez, M. X., Román-Rodríguez, M., Roca-Pardiñas, J., Zúñiga-Antón, M., García-Flaquer, A., Pericàs-Pulido, P., Sánchez-Recio, R., González-Álvarez, B., Rodríguez-Pastoriza, S., Gómez-Gómez, I., Motrico, E., Jiménez-Murillo, J. L., Rabanaque, I., & Clavería, A. (2022). *Analysis of the impact of social determinants and primary care morbidity on population health outcomes by combining big data: A research protocol. Frontiers in Medicine, 9*.
<https://doi.org/10.3389/fmed.2022.1012437>
- Estella Pérez, F. J., & Escobedo Ortega, N. (2024). *La inteligencia artificial en el sector salud: aplicaciones e impacto. I+S: Revista de la Sociedad Española de Informática y Salud, 158*, 21-24. <https://seis.es/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https://seis.es/wp-content/uploads/2024/03/IS-158.pdf>
- Gaceta Sanitaria. (2020). *Gac Sanit, 34*(1), Barcelona, ene./feb.
<https://doi.org/10.1016/j.gaceta.2018.07.015>
- García-Mayor, J., Moreno-Llamas, A., & De la Cruz-Sánchez, E. (2023). *Cómo el estatus socioeconómico afecta el estado de peso a través de estilos de vida relacionados con la salud: un análisis de clases latentes. European Journal of Cardiovascular Nursing, 22*(7), 730–744. <https://doi.org/10.1093/eurjcn/zvac101>
- Houghton, N., Bascolo, E., & Riego, A. del. (2020). *Socioeconomic inequalities in access barriers to seeking health services in four Latin American countries. Pan American Journal of Public Health, 44*, 1–9. <https://doi.org/10.26633/RPSP.2020.11>

- López-Mendizábal, L., Varea, C., Berlanga, A., Patricio, M. A., Molina, J. M., & Bartha, J. L. (2024). *Non-elective caesarean section risk assessment using machine learning techniques. Clínica e Investigación en Ginecología y Obstetricia, 51(3), 100949.*
<https://doi.org/10.1016/j.gine.2024.100949>
- Mora-Moreo, L., Estrada-Orozco, K., Espinosa, O., et al. (2023). *Caracterización de la población afiliada al sistema subsidiado de salud en Colombia: una revisión sistemática y metaanálisis. International Journal for Equity in Health, 22, 28.*
<https://doi.org/10.1186/s12939-022-01818-x>
- Negrín Hernández, M. Á. (2024). *Acceso a la atención sanitaria mediante seguros de salud comunitarios entre migrantes de Senegal.*
<https://www.scielosp.org/article/gs/2022.v36n5/409-415/>
- Organización Mundial de la Salud (OMS). (2019). *Desigualdades en salud: Un desafío global.*
<https://www.who.int/data/gho/publications/world-health-statistics>
- Ortiz-Ruiz, N. (2021). *Análisis de desigualdades en salud. Una propuesta cualitativa.*
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-386X2021000100006
- Pouye, R., Diallo, M. A., Nana, J. T., & Fall, N. (2024). *Effect of the policy of free health care for children under five on child undernutrition and social inequalities in health care use in Senegal.* <https://ojs.ual.es/ojs/index.php/eea/article/view/9781>
- Sánchez-de-Madariaga, R., Martínez-Romo, J., & Escribano, J. M. C. (2022). *Aprendizaje incremental semisupervisado con pocos ejemplos para descubrir reglas de asociaciones médicas. BMC Medical Informatics and Decision Making, 22, 20.*
<https://doi.org/10.1186/s12911-022-01755-3>

- Scheffer, I. E., & Liao, J. (2020). *Deciphering the concepts behind “Epileptic encephalopathy” and “Developmental and epileptic encephalopathy.” European Journal of Paediatric Neurology, 24(1), 11–14.* <https://doi.org/10.1016/j.ejpn.2019.12.023>
- Sepúlveda Plata, M. C., López Romero, L. A., & González, S. B. (2021). *Cumplimiento de la lista de verificación de seguridad de la cirugía en un hospital de Santander: Un estudio de corte transversal. Revista Cuidarte, 12(3), 1–17.* <https://doi.org/10.15649/cuidarte.2122>
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). *How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. Annual Review of Psychology, 1–22.* <https://doi.org/10.1146/annurev-psych-010418-102803>
- Sujatt, J. A. (2024). *La planificación de políticas de salud en una provincia periférica: El plan de salud de Chaco y su contexto social desde 2020.* <https://www.cuestionessociologia.fahce.unlp.edu.ar/article/view/CSe166>
- TerriData :: DNP. (s/f). Gov.co. Recuperado el 17 de octubre de 2025, de <https://terridata.dnp.gov.co/index-app.html#/descargas>
- Health Nutrition and Population Statistics. (s/f). Worldbank.org. Recuperado el 17 de octubre de 2025, de <https://databank.worldbank.org/source/health-nutrition-and-population-statistics>
- geoBoundaries. (s/f). Geoboundaries.org. Recuperado el 17 de octubre de 2025, de <https://www.geoboundaries.org/>
- Global burden of disease (GBD). (s/f). Institute for Health Metrics and Evaluation. Recuperado el 17 de octubre de 2025, de <https://www.healthdata.org/research-analysis/gbd>

Poverty and inequality platform. (s/f). Worldbank.org. Recuperado el 20 de noviembre de 2025,
de <https://pip.worldbank.org/country-profiles/IDN>

Apéndices

Apéndice A

Cronograma de Actividades

Fase / Actividad	Descripción	Duración estimada	Responsable / Herramienta
1. Definición del problema	Identificación de las desigualdades sanitarias, selección de indicadores y fuentes de datos oficiales (DANE, Ministerio de Salud, Banco Mundial).	Semana 1	Investigador principal
2. Recolección de datos	Obtención y limpieza de bases de datos sanitarias nacionales; consolidación del dataset “dataset_desigualdad_salud_wide_con_idas.csv”.	Semanas 2–3	Python (Pandas, GeoPandas)
3. Procesamiento y normalización	Integración de variables socioeconómicas, geográficas y demográficas; cálculo del IDAS (Índice de Desigualdad en Atención Sanitaria).	Semanas 4–5	Python, NumPy, Scikit-learn
4. Análisis descriptivo y correlacional	Exploración de tendencias, correlaciones y desigualdades regionales; generación de gráficos exploratorios.	Semanas 6–7	Plotly, Dash
5. Desarrollo del modelo predictivo	Implementación de regresión lineal y Random Forest para estimar IDAS en periodos 2022–2023.	Semanas 8–9	Scikit-learn
6. Diseño del Dashboard interactivo	Creación del tablero con tres pestañas: visualización geográfica, análisis estadístico y modelo predictivo.	Semanas 10–11	Dash + Bootstrap
7. Validación y visualización final	Evaluación de desempeño del modelo (R^2 y MAE), proyección del IDAS, generación de informes y gráficas.	Semana 12	Jupyter Notebook

8. Elaboración del informe final y presentación	Redacción técnica según normas APA, preparación de diapositivas y defensa oral.	Semanas 13–14	Word, PowerPoint
---	---	---------------	------------------

Apéndice B

Instrumentos de Recolección de Información

Instrumento / Fuente	Descripción del contenido	Frecuencia / Período de uso	Responsable de registro
Bases de datos del DANE	Datos censales y socioeconómicos a nivel departamental y municipal.	2005–2023	DANE / Investigador principal
Ministerio de Salud y Protección Social (SISPRO)	Indicadores de salud pública, mortalidad infantil, vacunación, afiliación al SGSSS.	Anual	MSPS / Observatorio Nacional de Salud
GeoBoundaries (Harvard CGA)	Límites geográficos administrativos (departamentos y municipios) en formato GeoJSON.	Última versión 2024	Harvard Dataverse
Banco Mundial / OMS (WHO Global Health Data)	Indicadores internacionales comparativos de acceso a salud.	2000–2023	Banco Mundial / OMS
Python Dash - Formulario de control	Interfaz digital para registro y actualización automática de variables sanitarias.	Durante ejecución del proyecto	Equipo técnico / Investigador principal