

**Seminario de investigación**

**Mario Andrés Hernández Moreno**

**Actividad 3: Formulación de anteproyecto**

**Dr. Oscar Cardona Morales**

**Universidad Autónoma de Manizales, Manizales**

**Especialización en Inteligencia Artificial**

**2024**

## 1. Planteamiento del problema

La Enfermedad Diarreica Aguda (EDA) representa un desafío significativo para la salud pública global, especialmente en regiones con recursos limitados y problemas de acceso a agua potable. En Ipiales, Nariño, una ciudad de aproximadamente 160.000 habitantes al suroccidente de Colombia, un reciente brote de EDA con 366 casos reportados en un período de seis días en septiembre de 2024 ha puesto en evidencia la urgente necesidad de implementar un sistema eficiente para la predicción y monitoreo de este tipo de enfermedades.

Este brote ocurre en un contexto de crisis hídrica agravada por una sequía prolongada, que ha comprometido la calidad del agua potable. Además, la contaminación del río Blanco, principal fuente de abastecimiento de la ciudad, junto con deficiencias en la infraestructura de tratamiento de agua, ha incrementado el riesgo de propagación de enfermedades.

La necesidad principal identificada es el desarrollo de un sistema de alerta temprana, predicción y monitoreo continuo para brotes de EDA impulsado mediante inteligencia artificial. Actualmente, las autoridades dependen de medidas correctivas de respuesta estandarizadas, como encuestas y análisis de muestras, lo que limita la capacidad de respuesta rápida y resulta insuficiente para prevenir nuevos brotes a tiempo.

El estado del arte en vigilancia epidemiológica y salud pública digital sugiere que la implementación de sistemas basados en Inteligencia Artificial (IA), especialmente con enfoques de aprendizaje automático (ML), puede mejorar significativamente la detección temprana y la respuesta a brotes de enfermedades infecciosas. Estudios previos han demostrado que modelos predictivos como redes neuronales artificiales (ANN), bosques aleatorios (Random Forest), árboles de decisión (DT), máquinas de soporte vectorial (SVM) e incluso técnicas de interpretabilidad como XAI (eXplainable Artificial Intelligence), pueden detectar patrones importantes y decisivos en grandes volúmenes de datos para determinar picos de enfermedades o tendencias de ciertas patologías o comportamientos.

El reto técnico principal implica el desarrollo e implementación de un sistema de IA que integre múltiples componentes, fuentes de datos y algoritmos para abordar la complejidad del problema. Esto incluye la obtención de datos precisos y en tiempo real, el preprocesamiento y garantía de calidad de los datos, la selección adecuada de variables o características, la implementación de modelos predictivos de aprendizaje automático, la integración y visualización de datos, la garantía de privacidad y seguridad en el manejo de la información y finalmente la educación y capacitación hacia el personal de salud en la toma de decisiones basada en datos.

## **2. Objetivos**

### **Objetivo general:**

Desarrollar e implementar un sistema de inteligencia artificial para la predicción y monitoreo de brotes de Enfermedad Diarreica Aguda (EDA) en Ipiales, Nariño, que permita una detección temprana y una respuesta más eficiente por parte de las autoridades sanitarias.

### **Objetivos específicos:**

1. Diseñar y construir un sistema para la recolección y procesamiento de datos en tiempo real que integre información epidemiológica, ambiental y socioeconómica importante para la predicción de brotes de EDA.
2. Desarrollar y validar modelos de aprendizaje automático con la capacidad de predecir con alta precisión la probabilidad de EDA en diferentes zonas de la ciudad de Ipiales, utilizando técnicas como redes neuronales artificiales y bosques aleatorios.
3. Implementar un sistema de visualización, control, alerta temprana y evaluación que permita a las autoridades sanitarias locales interpretar fácilmente los resultados del modelo y tomar decisiones basadas en datos de manera oportuna.

## **3. Justificación**

El desarrollo e implementación de un sistema de inteligencia artificial para la predicción y monitoreo de Enfermedad Diarreica Aguda (EDA) en el municipio de Ipiales, Nariño, se alinea con varios Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas, así como con políticas nacionales, planes de salud y recomendaciones a nivel regional.

En el corto plazo, este proyecto contribuye directamente al ODS 3: Salud y Bienestar, al mejorar la capacidad de respuesta ante enfermedades transmitidas por el agua y reducir la mortalidad infantil asociada a la EDA, promoviendo así una vida saludable y con bienestar para toda la población. También se alinea con el ODS 6: Agua Limpia y Saneamiento, al promover una gestión más eficiente de los recursos hídricos y mejorar la calidad del agua mediante un monitoreo más efectivo.

Además, el proyecto se enmarca en las recomendaciones de la Misión de Sabios de Colombia, específicamente en el foco de "Ciencias de la Vida y de la Salud". La Misión va de la mano de el Plan Nacional de Ciencia, Tecnología e Innovación en Salud (PNCTeIS) y enfatiza la importancia de desarrollar y aplicar tecnologías avanzadas, como la inteligencia artificial, para abordar desafíos sociales y de salud pública, donde se ejemplifica cómo esta tecnología puede utilizarse para mejorar la toma de decisiones en el sector salud y optimizar la asignación de recursos, contribuyendo a la reducción de mortalidad infantil y al control de brotes, una meta importante para prevenir muertes evitables y cumplir las metas globales de salud, como la erradicación de enfermedades prevenibles.

A nivel nacional, el proyecto se alinea con el Plan Decenal de Salud Pública de Colombia 2022-2031, que destaca la importancia de fortalecer los sistemas de vigilancia epidemiológica y la capacidad de respuesta ante emergencias de salud pública, por lo que la implementación de este

sistema de IA en el municipio de Ipiales, a mediano y largo plazo podría servir como un proyecto piloto escalable a otras regiones del país, contribuyendo así a la modernización y gestión integral del riesgo en salud pública, apoyando la toma de decisiones para reducir la mortalidad evitable y mejorar las condiciones de vida y bienestar de los colombianos.

En el contexto regional, el proyecto aborda directamente las necesidades identificadas en el Plan de Desarrollo Municipal de Ipiales y el Plan Territorial de Salud 2024-2027, que enfatiza la importancia de mejorar la calidad del agua, trabajar en la prevención de enfermedades transmitidas por infecciones relacionadas con este líquido, disminuir el Riesgo de Calidad de Agua -IRCA y fortalecer la capacidad de respuesta ante emergencias sanitarias.

Desde la perspectiva de los beneficiarios, este proyecto ofrecerá un impacto positivo tanto para la población de Ipiales y sus alrededores como para las entidades o empresas financiadoras. Las comunidades locales se beneficiarán de una mayor calidad de vida y un acceso a agua más segura, autoridades locales y regionales con herramientas más eficaces para la toma de decisiones, mejor capacidad de planificación y respuesta ante emergencias de este tipo. Por otro lado, para el Ministerio de Salud y Protección Social se obtendrá un modelo replicable para otras regiones del país, potenciando la modernización del sistema de salud pública y para las empresas privadas o de tecnología involucradas se desarrollará experiencia en aplicaciones de IA en salud pública, abriendo oportunidades de negocio en un sector en crecimiento. Por último, para las instituciones académicas participantes o la comunidad científica, existirán oportunidades de investigación y publicaciones en el campo de la IA aplicada a la salud pública.

Este proyecto no solo aborda una necesidad urgente de salud pública en Ipiales, sino que también se alinea con objetivos de desarrollo sostenible globales, recomendaciones nacionales de innovación, políticas de salud pública y planes de desarrollo territoriales, por lo que su implementación exitosa podría sentar las bases para una transformación en la forma en que se abordan las enfermedades transmitidas por el agua en Colombia y otros países en desarrollo.

#### **4. Metodología**

Para el desarrollo e implementación del sistema de inteligencia artificial para la predicción y monitoreo de brotes de EDA, se propone una metodología basada en elementos de CRISP-DM (Cross-Industry Standard Process for Data Mining), adaptada específicamente para este tipo de proyectos, basándose en las siguientes fases:

1. Comprensión del problema y planificación
  - Definición detallada de los objetivos del proyecto y los criterios de éxito.
  - Identificación de stakeholders (población, autoridades, gobierno, equipo de ciencia de datos, proveedores, comunidades, ONG's) y establecimiento de canales de comunicación.
  - Evaluación de recursos disponibles y limitaciones.
  - Planificación inicial del proyecto y establecimiento de metas específicas.

2. Comprensión y adquisición de los datos
  - Identificación de fuentes de datos relevantes (fuentes hídricas, plantas de tratamiento, hospitales, centros de salud, estaciones meteorológicas).
  - Diseño e implementación de sistemas de recolección de datos en tiempo real.
  - Evaluación de la calidad y completitud de los datos disponibles.
  - Documentación de las características y limitaciones de los datos.
  - Instalación de sensores en fuentes de agua estratégicos.
3. Preparación de los datos y Feature Engineering
  - Limpieza y preprocesamiento de datos (manejo de valores faltantes, outliers, balanceo, codificación, anonimización, entre otros procesos de tratamiento).
  - Integración de diferentes fuentes de datos.
  - Selección y creación de características relevantes para el modelo.
  - Transformación de datos para su uso en modelos de Machine Learning.
4. Modelado
  - Selección de algoritmos de ML apropiados (ANN, Random Forest, SVM, etc.).
  - Entrenamiento y validación de modelos utilizando técnicas de validación cruzada.
  - Optimización de hiperparámetros mediante técnicas como búsqueda en cuadrícula o búsqueda aleatoria.
  - Implementación de técnicas de XAI para mejorar la interpretabilidad de los modelos.
5. Evaluación y validación
  - Evaluación cuidadosa del rendimiento del modelo utilizando métricas apropiadas.
  - Validación del modelo con datos de prueba independientes y en diferentes escenarios.
  - Evaluación del impacto potencial del modelo en el contexto real de salud pública.
  - Revisión y ajuste de los objetivos del proyecto según sea necesario.
6. Despliegue e integración
  - Desarrollo del tablero de control o dashboard de visualización interactiva.
  - Implementación del sistema en un entorno de producción.
  - Integración con sistemas existentes de salud pública locales.
  - Capacitación del personal en el uso del tablero de control y el sistema de alarma.
7. Monitoreo y mantenimiento
  - Establecimiento de procesos para el monitoreo continuo del rendimiento del sistema.
  - Implementación de mecanismos de retroalimentación para la mejora continua del modelo.
  - Actualización regular del modelo con nuevos datos.
  - Mantenimiento y actualización de la infraestructura tecnológica.

## 8. Evaluación del impacto

- Recolección de datos sobre el uso y efectividad del sistema.
- Análisis del impacto en la detección temprana y respuesta a brotes de EDA.
- Evaluación de la satisfacción de los usuarios y stakeholders.
- Identificación de áreas de mejora y oportunidades de expansión.

Esta metodología se implementará de manera iterativa, con sprints de 2-3 semanas, permitiendo una adaptación ágil a los desafíos, problemas y cambios que puedan surgir durante el desarrollo del proyecto. Se mantendrá una comunicación constante con los stakeholders y se realizarán revisiones regulares del progreso y los objetivos del proyecto, de esta manera nos aseguramos que el sistema esté operativo en el plazo estimado, mientras se monitorea y mejora su rendimiento de manera continua.

## 5. Cronograma

Actividad	Responsable	Entregable	Duración (semanas)
Comprensión, planteamiento y contextualización del problema	Líder/Gerente del proyecto, científico de datos, ingeniero de Machine Learning, epidemiólogo	Documentación e informe de objetivos, alcance del proyecto, establecimiento de aspectos legales y éticos, planificación y requerimientos	3
Comprensión y adquisición de los datos	Ingeniero de datos, epidemiólogo, Empresa De Obras Sanitarias De La Provincia De Obando, Proveedor de sensores	Instalación de sensores, informe de las distintas fuentes de los datos y plan o proceso de adquisición	6
Preparación de los datos y Feature Engineering	Ingeniero de datos, científico de datos	Conjunto de datos preprocesados o dataset y documentación del proceso	6
Modelado	Científico de datos, ingeniero de ML	Modelos de ML entrenados, evaluados y validados	8

Evaluación	Científico de datos, epidemiólogo, expertos en salud pública	Informe de evaluación de los modelos utilizados	3
Implementación y desarrollo del tablero de control	Ingeniero de software, analista de datos	Dashboard o sistema de visualización y alerta temprana	3
Integración y capacitación	Ingeniero de ML, desarrollador Full-Stack, especialista o personal entrenado para ejecutar la capacitación	Sistema integrado o desplegado con la documentación técnica respectiva y personal capacitado	4
Monitoreo y optimización	Ingeniero de ML, analista de datos	Informes de rendimiento, actualizaciones y mejoras	4
Documentación, evaluación del impacto y comunicación	Líder/Gerente del proyecto, epidemiólogo, personal de salud entrenado y capacitado	Documentación completa, informe del impacto, recomendaciones y presentación final	3
Duración total estimada del proyecto		<b>40 semanas (9 meses)</b>	

Nota: algunas de las actividades planteadas, pueden realizarse al mismo tiempo y de esta manera optimizar el tiempo total del proyecto. Como todo proceso de ciencia de datos real, este cronograma está sujeto a cambios basados en el progreso real y los desafíos encontrados durante la implementación.

## 7. Referencias

- World Health Organization: WHO. (2024, 7 marzo). Diarrhoeal disease. <https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease>
- Instituto Nacional de Salud de Colombia. (2024). Boletín Epidemiológico Semanal. <https://www.ins.gov.co/buscador-eventos/Paginas/Vista-Boletin-Epidemiologico.aspx>
- Ligda P, Mittas N, Kyzas GZ, Claerebout E, Sotiraki S. Machine learning and explainable artificial intelligence for the prevention of waterborne cryptosporidiosis and giardiasis. Water Res. 2024 Sep 15;262:122110. doi: 10.1016/j.watres.2024.122110. Epub 2024 Jul 22. PMID: 39042970.
- Hussain M, Cifci MA, Sehar T, Nabi S, Cheikhrouhou O, Maqsood H, Ibrahim M, Mohammad F. Machine learning based efficient prediction of positive cases of waterborne diseases. BMC Med Inform Decis Mak. 2023 Jan 18;23(1):11. doi: 10.1186/s12911-022-02092-1. PMID: 36653779; PMCID: PMC9848024.
- Pras A, Mamane H. Nowcasting of fecal coliform presence using an artificial neural network. Environ Pollut. 2023 Jun 1;326:121484. doi: 10.1016/j.envpol.2023.121484. Epub 2023 Mar 21. PMID: 36958657.
- Petrau, A., & Sagarduy, M. (2024). Computer-implemented method for analysing water quality (WO2024156901A1). European Patent Office. <https://patentscope.wipo.int>
- Jutla, A. S., Usmani, M., & Colwell, R. (2024). Predictive models for infectious diseases (US2024029894A1). United States Patent and Trademark Office. <https://worldwide.espacenet.com>
- Misión de Sabios Colombia. (2019). Colombia hacia una sociedad del conocimiento: Recomendaciones de la Misión Internacional de Sabios 2019. Ministerio de Ciencia, Tecnología e Innovación. [https://minciencias.gov.co/sites/default/files/libro\\_mision\\_de\\_sabios\\_digital\\_1\\_2\\_0.pdf](https://minciencias.gov.co/sites/default/files/libro_mision_de_sabios_digital_1_2_0.pdf)
- Naciones Unidas. (2022b, mayo 24). Objetivos y metas de desarrollo sostenible - Desarrollo Sostenible. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>
- Alcaldía de Ipiales. (2023). Plan de Desarrollo Municipal. 2024-2027: Gobierno del Pueblo. [https://ipialesnarino.micolombiadigital.gov.co/sites/ipialesnarino/content/files/001208/603\\_97\\_plan-de-desarrollo-gobierno-del-pueblo.pdf](https://ipialesnarino.micolombiadigital.gov.co/sites/ipialesnarino/content/files/001208/603_97_plan-de-desarrollo-gobierno-del-pueblo.pdf)
- Alcaldía de Ipiales. (2023). Plan Territorial de Salud PTS. 2024-2027: Ipiales territorio de progreso en salud. [https://ipialesnarino.micolombiadigital.gov.co/sites/ipialesnarino/content/files/001208/603\\_96\\_plan-territorial-de-salud--pts.pdf](https://ipialesnarino.micolombiadigital.gov.co/sites/ipialesnarino/content/files/001208/603_96_plan-territorial-de-salud--pts.pdf)