

Inteligencia artificial como alternativa para la calidad del agua



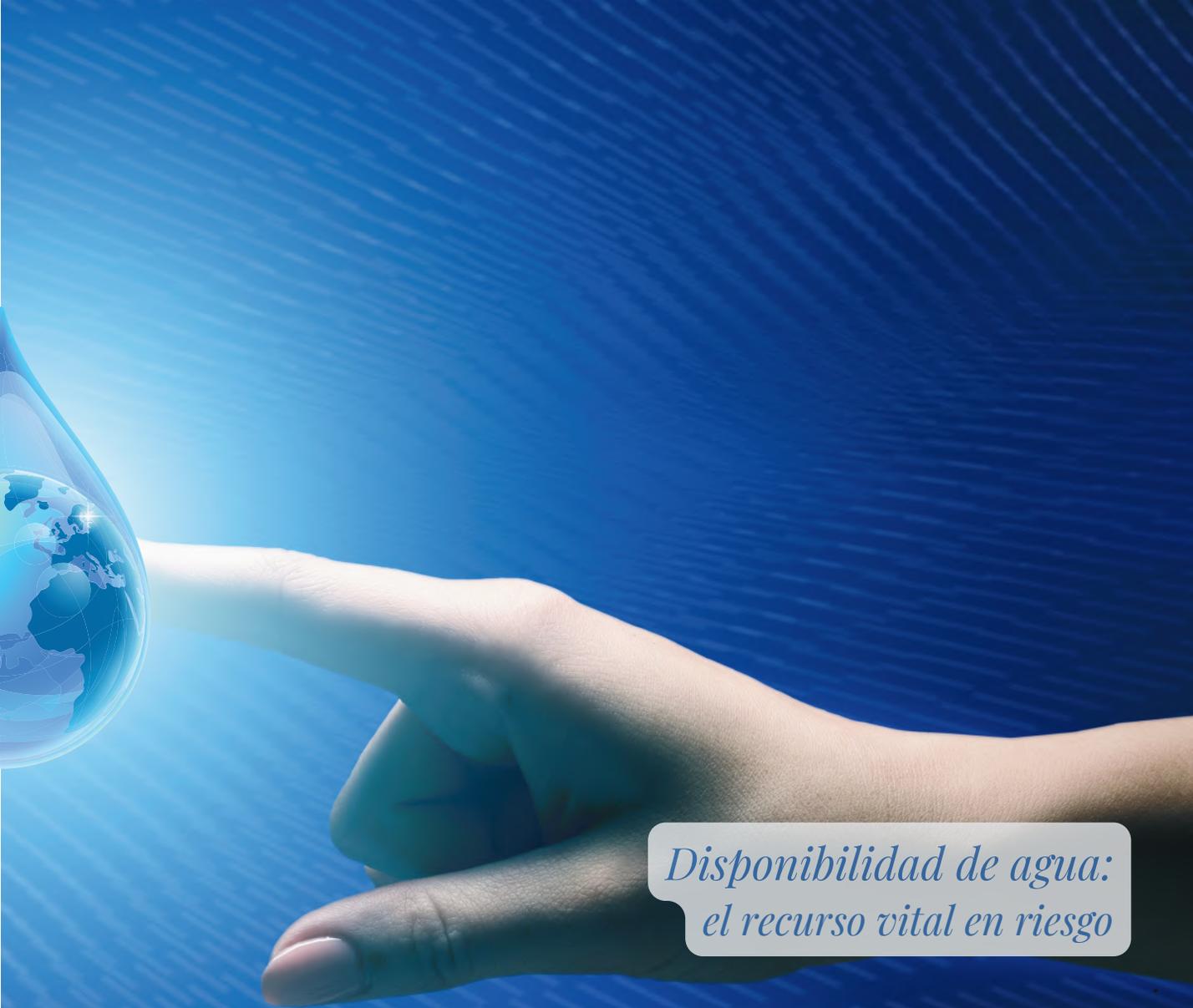
Damaris A. Zamudio Trujillo
Kimberly Mendivil García
Leonel E. Amábilis Sosa
José L. Medina Jimenez
Héctor Rodríguez Rangel

Artículo

Inteligencia artificial como alternativa para la calidad del agua

Cómo citar este artículo: Zamudio-Trujillo, DA, Mendivil-García K, Amábilis-Sosa LE, Medina-Jimenez JL, Rodríguez-Rangel H. 2024. Inteligencia Artificial como alternativa para la calidad del agua. Revista Ciencia y Naturaleza (1108).





*Disponibilidad de agua:
el recurso vital en riesgo*

El agua además de ser fundamental para la vida ha permitido el desarrollo económico y comercial de los países. Sin embargo, el crecimiento poblacional y la industrialización incrementan la demanda del líquido, generando la necesidad de desarrollar cada vez más infraestructura para abastecer este vital recurso, así como para el saneamiento de las aguas residuales generadas.



Actualmente sólo el 2.5% del agua en el planeta es técnicamente disponible para uso y consumo humano (1). Este vital líquido se encuentra accesible en lagos, ríos y depósitos subterráneos relativamente poco profundos (9-15 m) (2). Por lo que estos cuerpos de agua resultan de mayor importancia para el ser humano y su supervivencia.

En México, los ríos y arroyos representan una **red hidrográfica** de 633 mil km. Destacan 51 ríos por los que fluye el 87% del **agua superficial** y cuyas **cuencas** cubren el 65% de la superficie nacional (Figura 1). El 76% de agua dulce que se extrae de estos cuerpos de agua es para la actividad agrícola del país, y el 15% es para el abastecimiento público (1), por lo que es primordial conservar la calidad del agua superficial.



Figura 1. Principales cuencas y ríos de México.

El 76% de agua dulce que se extrae de ríos y presas es para la actividad agrícola del país.



Efecto de la actividad humana sobre la calidad del agua

La contaminación del agua en ríos y arroyos resulta una problemática de interés en gestión de recursos. Como se ha mencionado, los principales usos de estos cuerpos de agua, en México, son la agricultura y abastecimiento público. No obstante, dichos cuerpos superficiales sirven como depósitos de aguas residuales generadas en zonas urbanas, muchas de ellas sin recibir un tratamiento previo.



Lo anterior resulta en un problema serio para la calidad del agua de los cuerpos superficiales, debido a que las altas cantidades de contaminantes en el **agua residual** podría superar la capacidad de auto regeneración de estos cuerpos, lo que finalmente pone en riesgo a los ecosistemas y salud en general (3).

La calidad de los cuerpos de agua depende de características físicas y químicas como la turbidez, pH, presencia de bacterias y nutrientes. Estos últimos son un parámetro crítico debido a su influencia en el proceso de **eutrofización**. La eutrofización es generada por el aporte excesivo de nutrientes inorgánicos en los cuerpos de agua, principalmente nitrógeno y fósforo. Este exceso de nutrientes provoca el crecimiento de algas de manera acelerada (4).



Las proliferaciones de alga son llamadas floraciones (*algal bloom* en inglés) y el exceso de ellos deteriora la calidad del agua. Algunas floraciones de algas son perjudiciales para los humanos, ya que producen cantidades elevadas de toxinas, provocando enfermedades al entrar en contacto con ella o beberla (4). Estas alteraciones en la calidad del agua pueden ser causadas por factores naturales y por actividades humanas. La contaminación por nutrientes generalmente es consecuencia de actividades agrícolas y urbanas sin tratamiento eficiente.



En México, aproximadamente 21.6 millones de hectáreas del territorio nacional son utilizadas para siembra (2). El estado de Sinaloa es un ejemplo de los principales productores agrícolas en el país. Genera alrededor de 12,170 toneladas de alimentos, teniendo como principales cultivos el maíz y frijol. En el caso específico del maíz, requiere fertilización en una tasa de 200 kg de nitrógeno y 50 kg de fósforo por hectárea al año. Muchos de estos químicos son transportados a los cuerpos de agua por escorrentía en concentraciones muy altas, provocando procesos de eutrofización.

Existen estudios donde se evidencia el impacto de la actividad agrícola sobre la calidad del agua en la cuenca del río Culiacán, en Sinaloa (5). A través de herramientas de **sistemas de información geográfica** y estadística multivariada, el estudio determinó que, a medida que existe mayor porcentaje de suelo agrícola circundante a los sitios de monitoreo de la calidad del agua del río, las concentraciones de nutrientes (N y P) también iban incrementando. De esta manera se determinó que, en efecto, existe un impacto cuantificable de la actividad humana sobre la calidad del agua del río.



Necesidad de monitoreo de calidad del agua

Una mala calidad del agua limita el uso directo con humanos, dado que puede contribuir en transmisión de enfermedades. También, el medio ambiente sufre afectaciones, dado que una alta concentración de contaminantes puede provocar una baja concentración de oxígeno disuelto, lo que pone en riesgo los ecosistemas acuáticos (6). Estos dos casos y muchas otras afectaciones en la calidad del agua tienen su origen en las actividades antropogénicas propias de la urbanización y crecimiento de las poblaciones. Así, es indispensable monitorear la calidad del agua a fin de resguardar la salud pública y del medio ambiente antes de que los daños ocurran.

Actualmente existen normativas que establecen la composición física y química adecuada del agua según su uso. Esto nos permite asegurar que el agua utilizada para las actividades domésticas o agrícolas cumplen con los requisitos mínimos necesarios. Un ejemplo es la NOM-001-SEMARNAT-2021, la cual fue recientemente actualizada para adecuarse a las nuevas exigencias que se demandan en los vertimientos de agua residual en los cuerpos de agua nacionales.



Así mismo normativas como la NOM-003-SEMARNAT-1997 establece límites máximos permisibles de concentraciones en parámetros de calidad del agua en aguas residuales tratadas que permiten el reúso seguro en servicios públicos, misma que se encuentra en proceso de actualización.



Por otro lado, la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) también establece límites máximos permisibles de parámetros físicos, químicos y microbiológicos para el uso y consumo humano (7). Por su parte la organización mundial de la salud (OMS) elabora directrices sobre la calidad del agua ya sea para uso recreativo o consumo humano.

“Puede haber agua por todas partes y ni una gota para beber”
- Samuel Coleridge

Para asegurar el cumplimiento de las normativas de calidad del agua, se realizan intensivos monitoreos en los cuerpos de agua, cuyo análisis debe considerar el comportamiento de los parámetros por temporadas y/o zonas geográficas.



La Comisión Nacional del Agua (CONAGUA) es la instancia encargada de medir la calidad del agua de ríos, presas, lagunas y subterránea. La Red Nacional de Medición de la Calidad del Agua (RENAMECA) en 2022 contaba con 2,595 sitios de monitoreo en los diferentes cuerpos de agua y se utilizan más de 20 parámetros, en cada uno de ellos, para reportar la calidad del agua, entre los que destacan la demanda bioquímica del oxígeno total (DBO_TOT), la demanda química de oxígeno total (DQO_TOT), metales pesados (p.e. cadmio, plomo, arsénico, cromo), sólidos suspendidos totales (SST), entre otros parámetros fisicoquímicos y microbiológicos.



Los registros se generan a través de monitoreos bimestrales en cada una de las 757 cuencas hidrológicas del país. Se tienen en las bases de datos información histórica, con los que se realizan evaluaciones que permitan definir la medida de cumplimiento de los límites máximos permisibles de los parámetros de calidad del agua designados por CONAGUA para la utilización del recurso en las diferentes actividades requeridas (Figura 2).

Sin embargo, el trabajo desde la toma de muestra, transporte, determinaciones analíticas y exposición de resultados requiere contar con bastantes recursos económicos y humanos. Aunado los recursos necesarios para la adquisición de los datos, el manejo de estas bases de datos se vuelve compleja por lo que se requiera de herramientas que faciliten tanto el proceso como la interpretación de la información generada.

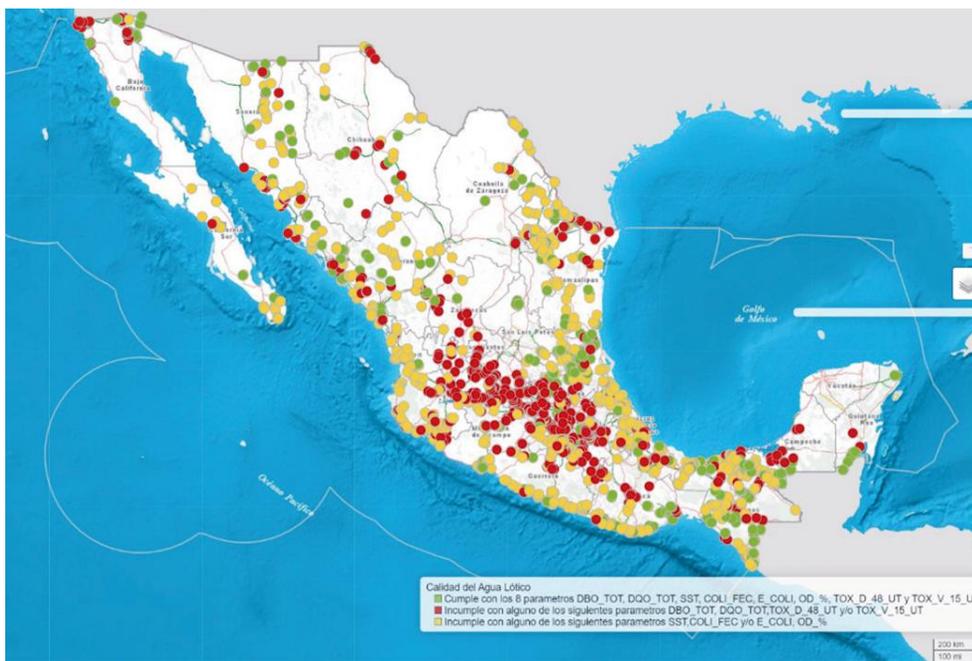
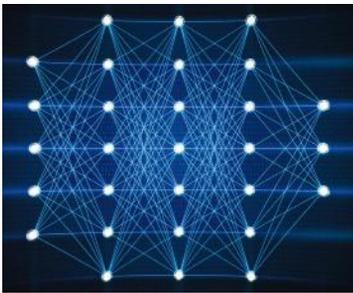


Figura 2. Distribución de sitios de monitoreo en cuerpos de agua superficiales.

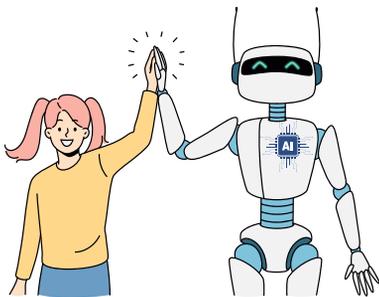


Alternativas tecnológicas para el monitoreo de la calidad del agua

La **Inteligencia Artificial** (IA) se centra en la creación de modelos que pueden mostrar comportamientos inteligentes semejantes al razonamiento humano con la capacidad de tomar decisiones, analizar datos en grandes cantidades, identificar patrones y tendencias, y formular predicciones de forma automática, con rapidez y precisión (8).



Existen modelos estadísticos como árboles aleatorios, máquinas de soporte vectoriales, *gradient boosting*, hasta redes neuronales artificiales, de las cuales, las redes perceptrones multicapa (MLP) permiten hacer predicciones de parámetros de calidad del agua. Para poder hacer uso de estos modelos se necesita de vasta información como datos de entrada como los datos de calidad del agua. Cada parámetro que se mide de calidad de agua sirve como una característica de entrada que puede ayudar a describir el comportamiento de la característica de salida, o un parámetro que quisiéramos predecir.



El **algoritmo de Random Forest** surge como la agrupación de varios árboles de clasificación; básicamente selecciona de manera aleatoria una cantidad de variables con las cuales se construye cada uno de los árboles individuales, y se realizan predicciones con estas variables que posteriormente serán ponderadas a través del cálculo de la clase más votada de los árboles que se generaron, para finalmente hacer la predicción del parámetro deseado (9).



En la Figura 3, se observa el procedimiento de predicción de parámetros de calidad del agua por medio del método Random Forest consta de tres etapas. En una primera etapa se seleccionan las variables que se utilizarán para predecir a la variable objetivo, en este caso se utilizan parámetros de calidad del agua para poder predecir la cantidad de algún otro parámetro presente en el agua. Para esto el algoritmo crea arboles de decisiones donde cada árbol se entrena con una parte aleatoria de los datos y toma decisiones basadas en características importantes.

Al ingresar nuevos datos, se le solicita a cada árbol que tome una decisión. Cada árbol vota por lo que considera que es la respuesta correcta. Finalmente, la decisión se toma con base en la mayoría de los votos de todos los árboles. Si más árboles votan por una decisión particular, esa es la que se elige. Es así como Random Forest puede brindar resultados más precisos, ya que considera múltiples decisiones tomadas por los conjuntos de árboles y decide basado en la mayoría.

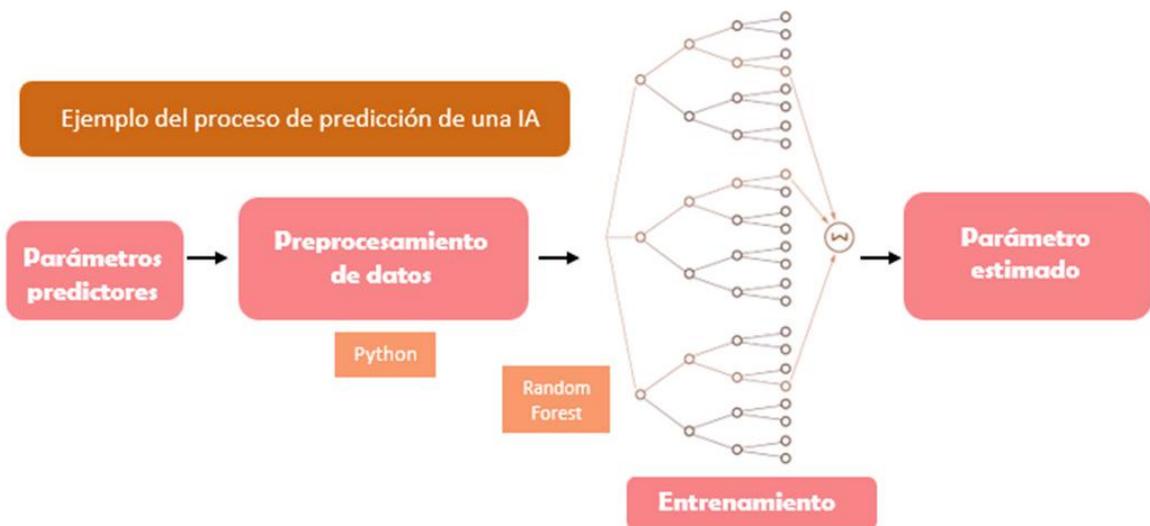


Figura 3. Diagrama del proceso de predicción de una IA



En la literatura especializada, la IA ha ganado relevancia debido a la rapidez con la que se obtienen los resultados con el mínimo requerimiento de recursos. Algunos campos de aplicación son la predicción de parámetros fisicoquímicos y microbiológicos, de temperaturas y en la reconstrucción de datos. Estos últimos para los casos en los que no es posible realizar actividades de muestreo, y evitar disminuir la precisión de los resultados calculados. Se tiene el ejemplo del estudio realizado en 2020 en “Red neuronal artificial para estimar un índice de calidad de agua” donde hacen uso de una red neuronal artificial (RNA) para emular el sistema neuronal biológico del procesamiento de información con el propósito de predecir parámetros fisicoquímicos y microbiológicos que influyen en la calidad del agua (10).

En regiones altamente influenciadas por actividades antropogénicas, la IA brinda una alternativa para la detección temprana de cambios en la composición del agua superficial a través de estimaciones de los parámetros de interés, detección de posibles fuentes puntuales de contaminación y generación de propuestas de mitigación de impacto ambiental.



Actualmente, en Sinaloa se están realizando las primeras pruebas con modelos de predicción de nutrientes a partir de otros parámetros de las bases de datos. Principalmente son de interés, conocer el comportamiento del nitrógeno y fósforo dado que la región es principalmente agrícola y el agua de los principales ríos y presas son indispensables para continuar la producción de alimentos y abastecimiento de la población.

Se utilizó la base de datos de la red de monitoreo de calidad del agua de la Comisión Nacional del Agua de la cuenca hidrológica "Río Presidio" (11). El conjunto de datos fue dividido en época de lluvia y estiaje. Para el procesamiento de los datos fue necesaria la reconstrucción de la base de datos debido que existían datos faltantes, para después proceder a la normalización estadística de los datos como se muestra en la Figura 4. Dicha normalización permite que todos los datos se encuentren entre rangos comparables, en este caso entre 0 y 1 para que de esta manera el algoritmo no priorice unas variables sobre otras basadas en su magnitud. Esto fue aplicado a todas las variables predictoras.

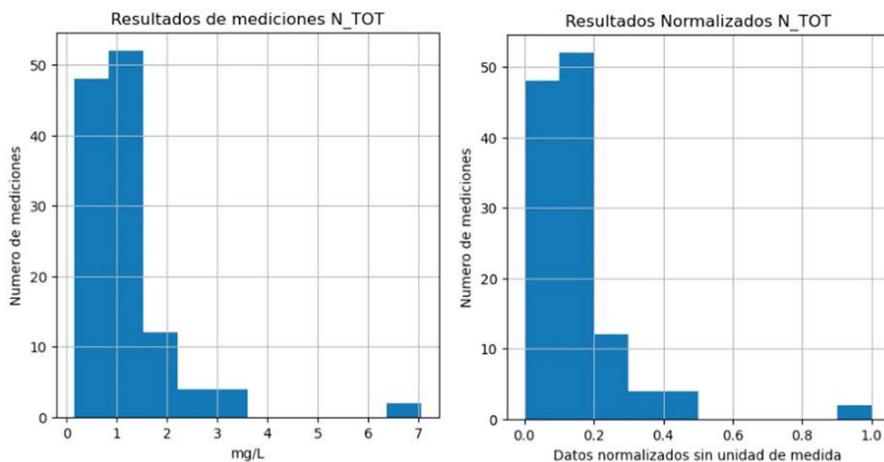


Figura 4. Comparativa de los datos normalizados y no normalizados de mediciones de nitrógeno en el río presidio desde 2012 hasta 2021.

Después, se identificaron los parámetros de calidad de agua que podrían servir para predecir otros parámetros específicos haciendo uso del modelo, como la demanda química de oxígeno, temperatura, pH, oxígeno disuelto, entre otras. Para el entrenamiento del sistema se utilizó el 80% de los datos y el 20% para validar si los resultados eran confiables. Para determinar la eficiencia del modelo se utilizaron como métricas de evaluación el coeficiente de determinación (R^2), la raíz del error cuadrático medio (RMSE) y la desviación absoluta media (MAE).



Como resultado se obtuvo un modelo capaz de predecir la concentración de nitrógeno que presenta un nivel de eficiencia alto con base en la métrica de coeficiente de determinación ($R^2 > 0.8$), con la posibilidad de seguir entrenando la inteligencia artificial y aumentar ese nivel de eficiencia. Lo que ofrece una alternativa al monitoreo convencional para realizar acciones correctivas ante los cambios en la calidad del agua.

En la Figura 5, se puede apreciar una gráfica donde se comparan los valores estimados por la IA contra los valores reales del modelo de predicción de nitrógeno en temporada de lluvia, el comportamiento de los datos nos indica que el modelo posee una buena capacidad predictora, llegando a tener un 87% de aciertos en sus predicciones, por lo que el modelo logra con el objetivo inicial de predecir las cantidades de nitrógeno presentes en el cuerpo de agua.

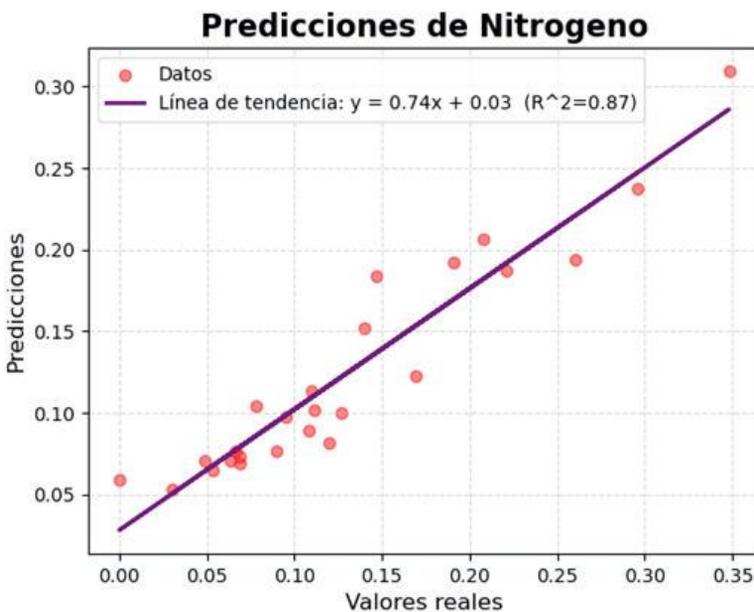


Figura 5. Representación gráfica del comportamiento de la predicción del nitrógeno con el modelo entrenado



Para llevar

La IA en el monitoreo de agua es un ejemplo de cómo se está revolucionando la gestión de la calidad del agua. Esta tecnología ayuda a mejorar la identificación de fuentes puntuales y mitigar la contaminación. A medida que la tecnología sigue evolucionando, el potencial de estos sistemas marca una diferencia real en la gestión de recursos hídricos, que permite la estimación de parámetros de calidad del agua como, en este caso, los nutrientes. A través de estos análisis es posible identificar cuando se dará un cuerpo de agua eutrofizado y actuar de manera pertinente de forma temprana. 🍀

Conceptos

Agua superficial: Es aquella que corre por cauces o se mantiene en embalses sobre la superficie del planeta. En un territorio particular, el agua superficial proviene fundamentalmente de la precipitación, aunque en algunos casos se le suma la de ríos que viajan desde otras regiones

Agua residual: Aguas procedentes de procesos productivos o del consumo humano. Previo tratamiento puede convertirse en aguas regeneradas susceptibles de reutilización si alcanzan los criterios de calidad adecuados para cada tipo de uso.

Algoritmo: Es una serie de pasos organizados, que describe el proceso que se debe seguir, para dar solución a un problema específico.

Coefficiente de determinación: Estima el porcentaje de explicación de la variable dependiente a partir de una o más variables independientes.

Cuenca: Territorio cuyas aguas fluyen todas hacia un mismo río, lago o mar; y a esta clase de cuencas se les llama "cuencas hidrográficas. Estas, son zonas de la superficie terrestre en donde (si fuera impermeable) las gotas de lluvia que caen sobre ella tienden a ser drenadas por el sistema de corrientes hacia un mismo punto de salida. Es decir, es una especie de "embudo" del territorio por el que escurre el agua desde las partes altas, hasta llegar a un punto en común, de donde sale el agua que fluye hacia otro lado.



Conceptos

Desviación absoluta media: Es una métrica utilizada para evaluar la precisión de un modelo de regresión. Suma las varianzas absolutas y divide el resultado entre el número de registros. La desviación absoluta media es una estadística de error que calcula el promedio de la distancia ente cada par de puntos de datos reales y ajustados.

Escorrentía: Agua de lluvia que discurre por la superficie de un terreno.

Eutrofización: La eutrofización es el excesivo crecimiento algal derivado de un desmedido aporte de nutrientes inorgánicos en los cuerpos de agua superficiales.

Inteligencia Artificial: Campo de la informática que se enfoca en el desarrollo de sistemas y tecnologías que pueden realizar tareas que normalmente requieren inteligencia humana.

Métricas de evaluación: Son herramientas estadísticas utilizadas para medir el rendimiento y la precisión de un modelo predictivo.

Monitorear: Observar mediante aparatos especiales el curso de uno o varios parámetros fisiológicos o de otra naturaleza para detectar posibles anomalías.

Oxígeno disuelto: es la cantidad de oxígeno gaseoso que esta disuelto en el agua, es fundamental para el desarrollo de vida de los peces, plantas, algas, y otros organismos.

Raíz del error cuadrático medio: Es una métrica que mide la cantidad de error entre dos conjuntos de datos. Se utilizó para expresar el error en la misma unidad que la variable dependiente original. En otras palabras, compara un valor predicho y un valor observado o conocido.

Random Forest: Algoritmo de aprendizaje automático que combina múltiples árboles de decisión individuales, cada uno entrenado en una muestra aleatoria del conjunto de datos de entrenamiento y utilizando diferentes subconjuntos de características.

Red hidrográfica: Una red hidrográfica es un sistema de circulación lineal, jerarquizado y estructurado que asegura el drenaje de una cuenca; específicamente una cuenca hidrográfica.

Sistemas de información geográfica: Un software específico que permite a los usuarios crear consultas interactivas, integrar, analizar y representar de una forma eficiente cualquier tipo de información geográfica referenciada asociada a un territorio, conectando mapas con bases de datos.

Para Consulta

-  IMCO. 2023. Centro de Investigación en Política Pública. 2023. [[Link](#)]
-  SIAP. 2021. Anuario Estadístico de la Producción Agrícola. 2021. [[Link](#)]
-  Pérez B, Javier C. 2022 Reutilización de aguas depuradas en la industria. Universidad Distrital Francisco José de Caldas. [[Link](#)]
-  Romero GM. 2020. El Proceso de Eutrofización: otra de las múltiples caras de la contaminación. MoleQla: revista de Ciencias de la Universidad Pablo de Olavide 33-36.
-  Mendivil K, Amábilis LE, Rodríguez AE, *et al.* 2020. Assessment of intensive agriculture on water quality in the Culiacan River basin, Sinaloa, Mexico. Environmental Science and Pollution Research 27(23): 28636-28648.
-  Calla-Navarro JA. 2019. Actividades Antrópicas y calidad del agua. UNIVERSIDAD NACIONAL DE CAJAMARCA. [[Link](#)]
-  FAO. 2014. Organización de las naciones unidas para la alimentación y agricultura. [[Link](#)]
-  Porcelli AM. 2020. La inteligencia artificial y la robótica: sus dilemas sociales, éticos y jurídicos. Derecho Global, Estudios sobre derecho y justicia. [[Link](#)]
-  Medina-Merino RF, Ñique-Chacón C. 2017. Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. [[Link](#)]
-  Quiñones L. 2020. Red neuronal artificial para estimar un índice de calidad de agua. [[Link](#)]

Dra. Nuvia Katún Moreno

Editor Asociado Revista CyN

Diseño de publicación: Yareli Fiburcio

Crédito de imágenes en orden de aparición: Blue Planet Studio (Getty Images, GI), Prostock-Studio (GI), Geolimages, Aleksandr Slobodanyk (Pexels, P), Free Nature Stock (P), Drawlab19, andipantz (GI), CasarsaGuru (GI), Trixcis (GI), sedmak (GI), icojam, Chinnapong (GI), Vintage Illustrations, nemke (GI), Sergei Chuyko (GI), DuxX (GI), Galda Picture, Vertigo3d (GI), Igor Kutyaev (GI), Miftah, leMBERGvector, Chatreep (GI), Andrzej Boldaniuk (GI), 4X-image, Felicia Smith, phokin.



Damaris A. Zamudio Trujillo

Instituto Tecnológico Nacional de México Campus Culiacán. Egresada de la carrera Ingeniería Ambiental con especialidad en calidad del agua, y diplomado en agua y saneamiento. Reciente participación oral en el XLV encuentro nacional de la Academia Mexicana de Investigación y Docencia en Ingeniería Química.

contacto: damarizamudio27@gmail.com



Kimberly Mendivil García

CONAHCYT-Tecnológico Nacional de México Campus Culiacán. Investigadora Posdoctorante por México, miembro del SNII, nivel 1, y de la Red Temática de Gestión de la Calidad y Disponibilidad del Agua. Su especialidad es el estudio de la calidad del agua y el efecto del uso de suelo en cuencas hidrológicas.

contacto: kimberly.mg@culiacan.tecnm.mx



Leonel E. Amábilis Sosa

CONAHCYT-Tecnológico Nacional de México Campus Culiacán. Es investigador por México, miembro de la Red Temática de Gestión de la Calidad y Disponibilidad del Agua y de la International Water Association. Su especialidad es la modelación y análisis de la generación, transporte y destinos de contaminantes, su mitigación y control mediante soluciones basadas en la naturaleza.

contacto: leonel.as@culiacan.tecnm.mx



José Luis Medina Jimenez

División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México Campus Culiacán. Actualmente es estudiante del Doctorado en Ciencias de la Ingeniería, desarrollando el tema de Estimación y simulación de parámetros de la calidad del agua en cuerpos superficiales de Sinaloa implementando inteligencia artificial.

contacto: jose.mj@culiacan.tecnm.mx



Héctor Rodríguez Rangel

División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México Campus Culiacán. Actualmente es Profesor Titular C y miembro del Sistema Nacional de Investigadoras e Investigadores, nivel 1. Sus líneas de investigación se centran en optimización, computo inteligente y reconocimiento de patrones.

contacto: hector.rr@culiacan.tecnm.mx