

TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Culiacán



TITULACIÓN INTEGRAL

Opción:

Tesis Profesional

“Estimación de parámetros de calidad del agua en cuenca hidrológica
Río presidio con el apoyo de algoritmos de inteligencia artificial”

ALUMNO(A):

Damaris Alejandra Zamudio Trujillo

CARRERA:

Ingeniería Ambiental

No. DE CONTROL:

19171274

ASESOR(A):

Dra. Kimberly Mendívil García

Culiacán, Sinaloa

HOJA DE AUTORIZACIÓN



Educación
Secretaría de Educación Pública



Instituto Tecnológico de Culiacán

DIVISION DE ESTUDIOS PROFESIONALES
ASUNTO: **Autorización de Impresión**

Culiacán, Sinaloa, 11/03/2025 9:37

C. DAMARIS ALEJANDRA ZAMUDIO TRUJILLO
PASANTE(S) DE LA CARRERA DE INGENIERÍA AMBIENTAL
P R E S E N T E.

De acuerdo con fallo emitido por la Comisión Revisora Integrada por:

Cédula Profesional

Presidente:	Doctorado en Ingeniería Ambiental Leonel Ernesto Amabilis Sosa	10359494
Secretario:	Doctorado en Ciencias en Ingeniería Eléctrica Opción en Sistemas Computacionales Héctor Rodríguez Rangel	9972835
Vocal:	Doctorado en Ciencias de la Ingeniería Kimberly Méndivil García	13179723

Y considerando que reúne los requisitos establecidos en la Opción: **TITULACIÓN INTEGRAL: TESIS PROFESIONAL "ESTIMACIÓN DE PARÁMETROS DE CALIDAD DE AGUA EN CUENCA HIDROLÓGICA RIO PRESIDIO CON EL APOYO DE ALGORITMOS DE INTELIGENCIA ARTIFICIAL"**, damos a usted nuestra autorización para que proceda a imprimir su trabajo profesional.

Sin otro particular, quedo de usted.

ATENTAMENTE
Excelencia en Educación Tecnológica®


LIC. CONCEPCIÓN MENDOZA ROSALES
JEFA DE LA DIVISIÓN DE ESTUDIOS PROFESIONALES

 **EDUCACIÓN** 
SECRETARÍA DE EDUCACIÓN PÚBLICA INSTITUTO TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CULIACÁN
DEPARTAMENTO DE DIVISIÓN DE ESTUDIOS PROFESIONALES



Juan de Dios Báltiz 310 Pte. Col. Guadalupe C.P. 80220
Culiacán, Sinaloa. Tel. 667-454-0100
tecnm.mx www.culiacan.tecnm.mx



AGRADECIMIENTO

Me gustaría agradecer a mis papás Teresita de Jesús Trujillo Segoviano y Alejandro Zamudio Urquidez, por brindarme todas las herramientas necesarias a lo largo de mi vida para hacer esto posible, por criarme con los mejores valores posibles y llenar nuestro hogar con mucho amor y cariño. También quiero agradecer a mis compañeros de clase por acompañarme y apoyarme en este camino, por los buenos momentos y el gran compañerismo que me ofrecieron durante esta etapa. A mis amistades mas largas cuya presencia en mi vida ha sido fundamental para terminar de formarme como persona y por estar siempre a mi lado, como mi segunda familia. Y a mi perro que casi me acompañó durante toda mi formación académica, pero logró terminar el camino a mi lado.

Y a mis profesores y asesores, quienes me transmitieron todos los conocimientos académicos para la elaboración de este trabajo e incluso de la vida misma, muchas gracias por la confianza y por todo el apoyo otorgado.

Estimación de parámetros de calidad del agua en cuenca hidrológica Río presidio con el apoyo de algoritmos de inteligencia artificial

RESUMEN

El monitoreo proporciona información crucial sobre el estado de la biodiversidad, la calidad del agua y la presencia de contaminantes, siendo fundamental para la conservación de hábitats acuáticos. Además, es de gran importancia para la toma de decisiones en la retribución de recursos y permite prevenir y gestionar desastres naturales, controlar la eutrofización, cumplir normativas ambientales y respaldar investigaciones científicas a largo plazo. En muchas ocasiones realizar el muestreo que se necesita para llevar a cabo un monitoreo de manera correcta, puede presentar problemas que dificultan su ejecución por lo que la aplicación de una herramienta como lo es la inteligencia artificial resulta muy útil, en este caso únicamente se estiman con su ayuda a los nutrientes considerados en la eutrofización de los cuerpos de agua, la cual es uno de los desafíos ambientales con mayor importancia en el recurso hídrico. Para esto se utilizó al algoritmo "RandomForest" que permitió entrenar 4 modelos, logrando estimar nitrógeno y fósforo durante la época de lluvia y de estiaje respectivamente. Dando como resultado modelos con capacidad predictiva teniendo coeficientes de determinación desde el 75% al 90% y con rango de mejoría. Además de contar con RMSE y MAE considerablemente bajos encontrándose en los rangos desde 0.035 al 0.064 y del 0.021 al 0.040 respectivamente.

Palabras clave: Nutrientes, Monitoreo, Inteligencia artificial, Eutrofización

ÍNDICE DE CONTENIDO

Contenido

1	INTRODUCCIÓN	1
1.1	Antecedentes	1
1.2	Descripción del problema.....	4
1.3	Hipótesis o preguntas de investigación	5
1.4	Objetivo general	6
1.5	Objetivos específicos	6
1.6	Justificación.....	6
1.7	Alcance.....	11
2	FUNDAMENTO TEÓRICO	13
2.1	Ríos y su temporalidad	13
2.1.1	Importancia social y económica de los ríos	13
2.1.2	Época de lluvia y estiaje en ríos	15
2.1.3	Impacto por construcción de presas.....	18
2.1.4	El recurso agua en México	19
2.2	Importancia de los nutrientes	21
2.2.1	Problemática del Nitrógeno en cuerpos de agua	21
2.2.2	Problemática del Fósforo en cuerpos de agua	23
2.2.3	Impactos ambientales de nutrientes en exceso	24
2.2.4	Monitoreo De Calidad De Agua	25
2.2.5	Red De Monitoreo De Conagua	26
2.3	Inteligencia Artificial	26
2.3.1	Inteligencia artificial y la predicción de datos.....	26

2.3.2	Aplicaciones de la ciencia de datos en calidad del agua	27
2.3.3	Modelos predictores.....	28
2.3.4	Métricas de evaluación	33
3	MATERIALES Y MÉTODO	36
3.1	Obtención y exploración de base de datos.....	36
3.2	Preprocesamiento de Datos.....	37
3.3	Selección de Características Predictoras	38
3.4	Entrenamiento de inteligencia artificial.....	39
3.5	Manejo de Métricas de Evaluación	39
3.5.1	Aplicación del coeficiente de determinación.....	39
3.5.2	Aplicación de la Raíz del Error Cuadrático Medio (RMSE)	40
3.5.3	Aplicación de la Desviación Absoluta Media	40
4	RESULTADOS	41
4.1	Evaluación del modelo	41
5	CONCLUSIONES.....	50
6	COMPETENCIAS DESARROLLADAS.....	52
7	FUENTES DE INFORMACIÓN.....	53
8	REGISTRO DE PRODUCTOS	60
8.1	Artículo publicado en revista de divulgación científica.....	60
8.2	Participación como expositora en congreso internacional.....	60
8.3	Presentación de cartel en el Primer Congreso de Investigadoras Sinaloenses ...	61

ÍNDICE DE FIGURAS

<i>Figura 1-1 Despliegue de los puntos de monitoreo en cuerpos de agua superficiales en México (CONAGUA, 2022).</i>	8
<i>Figura 1-2 Ubicación geográfica de Río presidio (CONSELVA, 2023).</i>	10
<i>Figura 1-3 Vista satelital de Río Presidio (Google Earth, 2022).</i>	12
<i>Figura 2-1 Carta de climas en Sinaloa usando la clasificación propuesta por la SEP. (INEGI, 2023).</i>	16
<i>Figura 2-2 Carta de climas del estado de Durango 1:1 000 000 (INEGI, 2019).</i>	17
<i>Figura 2-3 Volumen concesionado para usos consuntivos, 2019 (hm³) (CONAGUA, 2021).</i>	21
<i>Figura 4-1. Correlación entre parámetros usando la correlación Pearson en Python.</i>	42
<i>Figura 4-2: Visualización objetiva de los valores predictivos de Nitrogeno Total vs los valores reales en temporada de estiaje.</i>	43
<i>Figura 4-3: Visualización objetiva de los valores predictivos de Nitrogeno Total vs los valores reales en temporada de lluvia.</i>	44
<i>Figura 4-4: Resultado predictivo del Fosforo Total vs los valores reales en temporada de estiaje.</i>	45
<i>Figura 4-5: Resultado predictivo del Fosforo Total vs los valores reales en temporada de lluvia.</i>	46
<i>Figura 4-6: Criterios ecológicos de calidad del agua CE-CCA-001/89 (SEMARNAT, 2018)</i>	49
<i>Figura 8-1 Carta de aceptación de contribución a revista de divulgación “Ciencia y Naturaleza”</i>	60
<i>Figura 8-2. Reconocimiento por presentación de trabajo en congreso “AMIDIQ”</i>	61
<i>Figura 8-3. Carta de aceptación a presentación de cartel en el Primer Congreso de Investigadoras Sinaloenses y cartel presentado</i>	62

ÍNDICE DE TABLAS

TABLA 4-1 SEPARACIÓN DE CONJUNTO DE DATOS POR TEMPORADA Y SELECCIÓN DE CARACTERÍSTICAS.	41
TABLA 4-2: EVALUACIÓN DE MODELOS.....	46
TABLA 4-3: DIFERENCIA NUMÉRICA ENTRE VALORES REALES Y PREDICCIONES EN EL MODELO 2.	47

1 INTRODUCCIÓN

1.1 Antecedentes

El monitoreo de la calidad del agua consiste en la recopilación y el análisis sistemático de datos con el fin de identificar las variaciones temporales y espaciales en parámetros químicos, físicos y biológicos específicos. Este proceso permite evaluar si las características del agua son adecuadas para usos recreativos, consumo doméstico o aplicaciones industriales, así como determinar si es necesario implementar medidas para la protección de los ecosistemas acuáticos. (Chapman, 1996).

La salud de los cuerpos de agua se encuentra fuertemente relacionada con las características fisicoquímicas que presenten. Hay muchas maneras de monitorear la calidad del agua y se basan según los lineamientos y objetivos que persigue el programa de monitoreo planteado. Entre algunos métodos usuales se pueden distinguir tres tipos, Cuantificación de las características bióticas y abióticas que, en función del rubro también pueden ser de tipo físico, químico o biológico, o incluso cualitativo como el visual. (Aguas Urbanas, 2018). El monitoreo de variables fisicoquímicas considera parámetros como el color, sabor, olor, pH, turbidez, conductividad, niveles de amonio, etc. mientras que el método de monitoreo biológico considera la presencia de bacterias coliformes y el estudio de los organismos vivos que habiten en el cuerpo de agua, y se pueden utilizar para conocer la calidad del agua y en consecuencia estimar la salud del cuerpo de agua.

Existen diversos estudios en la literatura que se han enfocado en el análisis de la calidad del agua por medio de monitoreos a cuerpos de agua superficiales. Tal es el caso de Cañete (2019), en el cual el estudio se centra en aportar al control de la contaminación hídrica, enfocándose principalmente en la contaminación de los efluentes líquidos no tratados que contribuyen en la degradación y contaminación del río Paraguay. Esta investigación se llevó a cabo a través de actividades *in situ* como la toma de muestras y el análisis en el río. Entre los resultados se destaca la ausencia de un sistema o programa de supervisión constante que posibilite la evaluación de la calidad del agua del Río Paraguay destinada al consumo humano. El monitoreo llevó a concluir que el efluente cuenta con un grado altamente contaminante, fuera de los parámetros de vertido, de acuerdo con los valores permisibles establecidos según Resolución N° 222/02 “por la en el cual se establece el Padrón de Calidad de las Aguas Territorio Nacional”.

Por su parte, Larrea et al. (2022) aborda los principales aspectos relacionados con el monitoreo de la calidad de las aguas en ecosistemas dulceacuícolas superficiales. Relatan su importancia debido al aumento de la población, la expansión de las actividades agrícolas e industriales, además de la contaminación fecal producida por el ser humano. Además, se comparan ejemplos de investigaciones de diferentes partes del mundo donde se explican los aspectos para tener en cuenta durante un monitoreo de calidad de agua, especialmente el caso del río Almendares (La Habana, Cuba), el cual ha sido estudiado por un periodo de más de 15 años. Sugieren que para llevar a cabo un monitoreo es importante realizar un análisis al lugar donde se realizará el muestreo con el fin de facilitar la selección de los indicadores de

calidad del agua que se analizarán. Todo esto contribuye a mejorar el proceso de monitoreo, optimizar la utilización de recursos y minimizar los gastos asociados.

Finalmente, Carbajal y Sánchez (2017) llevaron a cabo un estudio empleando un nuevo modelo computacional en granjas de cultivo extensivo de camarón *Litopenaeus vannamei* con el objetivo de estimar la calidad del agua. El monitoreo fue realizado por el Instituto Tecnológico de Sonora (ITSON), en colaboración con el Centro de Investigaciones Biológicas del Noreste (CIBNOR), en la granja camaronera “Gez Acuícola”, ubicada en el ejido “Rancho Chapo”, Huatabampo, Sonora, México.

Para el estudio se seleccionaron tres parámetros clave: temperatura, oxígeno disuelto y salinidad. A través de una red neuronal artificial (RNA), se estableció un sistema de clasificación basado en conjuntos de concentraciones, permitiendo asociar cada conjunto a un estado de la calidad del agua y generar un indicador global con categorías como: excelente, bueno, regular y deficiente. Además, la RNA analiza y aprende las relaciones y dependencias entre los parámetros, lo que mejora la precisión de las evaluaciones realizadas. Los resultados indicaron que las redes neuronales artificiales constituyen una herramienta eficaz para identificar estados óptimos o deficientes de la calidad del agua en este tipo de cultivos.

De esta manera múltiples autores e investigadores han visto una oportunidad de apoyo viable en esta nueva herramienta inteligente que permite facilitar las tareas de análisis de datos en múltiples áreas, por lo que resulta viable pensar en un escenario donde resulte útil en términos de calidad de agua, sobre todo en el monitoreo de esta.

1.2 Descripción del problema

Los ríos son considerados ecosistemas altamente diversos y productivos, que aportan en gran medida al crecimiento económico, la seguridad alimentaria y el bienestar humano general. (UNEP, 2022). A lo largo de los siglos, los ríos y otros cuerpos de agua han sido aprovechados para diversos fines, como el suministro de agua a las ciudades, la generación de energía hidroeléctrica y el riego agrícola. Estas funciones se han facilitado mediante la construcción de grandes proyectos de infraestructura, los cuales han promovido el crecimiento, generado oportunidades y contribuido al bienestar económico de las comunidades. (Sukhdev, 2018).

El deterioro en el estado de salud de cualquier tipo de medio acuático dentro de la hidrología, tales como ríos, corrientes y arroyos, constituye un problema relevante en la gestión de los recursos hídricos. Estos cuerpos superficiales se utilizan principalmente para la agricultura y el abastecimiento público, pero también reciben aguas residuales provenientes de áreas urbanas, muchas de las cuales no cuentan con tratamiento previo.

Esta situación representa un riesgo significativo para la calidad del agua, ya que las altas concentraciones de contaminantes presentes en las aguas residuales pueden superar la capacidad de auto-regeneración de los cuerpos de agua, afectando tanto la salud pública como la de los ecosistemas.

La calidad del agua es un factor fundamental para el bienestar de los ríos y puede verse alterada por distintos factores, como el uso del suelo o el manejo de los vertimientos. Su evaluación se realiza mediante la medición de parámetros físicos,

químicos y biológicos, los cuales permiten identificar variaciones desde condiciones naturales hasta situaciones impactadas por actividades humanas o fenómenos naturales.

La Comisión Nacional del Agua (CONAGUA) está a cargo de realizar monitoreos en las cuencas hidrológicas del país, incluyendo la cuenta del río presidio, logrando recopilar una gran cantidad de datos respecto a las características que conforman los parámetros de calidad del agua. Estos monitoreos conllevan un gasto de recursos económicos, de personal calificado y de tiempo para realizar el muestro además de generar problemas de accesibilidad y de seguridad para el personal debido a la ubicación geográfica del río. Durante la emergencia sanitaria por CoVid-19 la realización de estos monitoreos presentó más dificultad debido a la limitación de recursos tanto económicos como humanos por lo que existe una escasez de datos en los años 2019, 2020 y 2021.

1.3 Hipótesis o preguntas de investigación

Hipótesis

- Es posible estimar los nutrientes de manera asertiva con el algoritmo de inteligencia artificial utilizado.
- Las concentraciones de nutrientes estimadas en la cuenca indican un nivel alto de eutrofización debido al tipo de actividad que se realiza actualmente en el suelo alrededor de la zona de estudio.

Pregunta de investigación

¿Qué implica la presencia de una herramienta como lo es la inteligencia artificial para los muestreadores?

1.4 Objetivo general

Comprobar la viabilidad del uso de la inteligencia artificial como una herramienta para el monitoreo de calidad del agua

1.5 Objetivos específicos

1. Identificar las concentraciones de nitrógeno y fósforo en la cuenca hidrológica río Presidio para la época de lluvia y estiaje.
2. Identificar los parámetros adecuados para utilizar como predictores de nitrógeno y fósforo en la cuenca del río presidio.
3. Estimar parámetros de nitrógeno y fósforo en la cuenca con el apoyo de inteligencia artificial.

1.6 Justificación

Durante siglos, los ríos se han utilizado para proporcionar una variedad de beneficios, incluidos agua urbana, energía hidroeléctrica e irrigación, a través de proyectos de infraestructura a gran escala que han impulsado el crecimiento, las oportunidades y las economías en todo el mundo (Sukhdev, 2018). Cuando un cuerpo de agua presenta una calidad deficiente restringe el aprovechamiento de los servicios ecosistémicos para los humanos. Un ejemplo claro es el fungir como vector para enfermedades y generar problemas de salud a largo plazo. De manera similar o menos directa para la salud, los niveles de oxígeno decaen ante la presencia de

contaminantes en los sistemas acuáticos. (Rodríguez-Lara, y otros, 2022). En gran parte este tipo de consecuencias son causadas por actividades antropogénicas las cuales solo han aumentado en los últimos años por el crecimiento de la población.

En México, la calidad del agua se regula mediante diversas normativas que establecen los parámetros físicos y químicos que deben cumplir según el uso previsto, asegurando que sea apta para actividades domésticas o agrícolas. Por ejemplo, la NOM-001-SEMARNAT-2021 sufrió marcadas actualizaciones relacionadas con los nuevos requisitos de ingeniería ambiental que ya consideran el vertimiento y/o contenido de compuestos recalcitrantes o tóxicos en los cuerpos de agua. (SEMARNAT, 2022). Estos cambios se alinean a los tratados internacionales de los que México forma parte recientemente. En efecto, los nuevos lineamientos de descarga de aguas residuales tratadas y no tratadas son muy similares a los de los países desarrollados tanto en Europa como en América.

Para garantizar que se cumplan las regulaciones sobre la calidad del agua, se llevan a cabo exhaustivos análisis en los cuerpos acuáticos, teniendo en cuenta el comportamiento de los indicadores según la temporada y/o la ubicación geográfica. La CONAGUA es la entidad responsable de evaluar la calidad del agua en ríos, embalses, lagunas y aguas subterráneas. a red oficial de monitoreo cuenta con 2,595 puntos de control distribuidos en diversos cuerpos de agua, en los cuales se analizan más de 20 variables para evaluar la calidad del recurso hídrico. Entre las principales destacan la demanda bioquímica de oxígeno (DBO), la demanda química de oxígeno (DQO), la presencia de metales pesados, así como diversos parámetros fisicoquímicos y microbiológicos.

Los registros se obtienen mediante monitoreos bimestrales realizados en cada una de las 757 cuencas hidrológicas del país. La información recopilada se integra en bases de datos históricas que permiten llevar a cabo evaluaciones orientadas a determinar el grado de cumplimiento de los límites máximos permisibles establecidos para el uso del agua en las distintas actividades productivas y sociales. (Véase Figura 1-1).

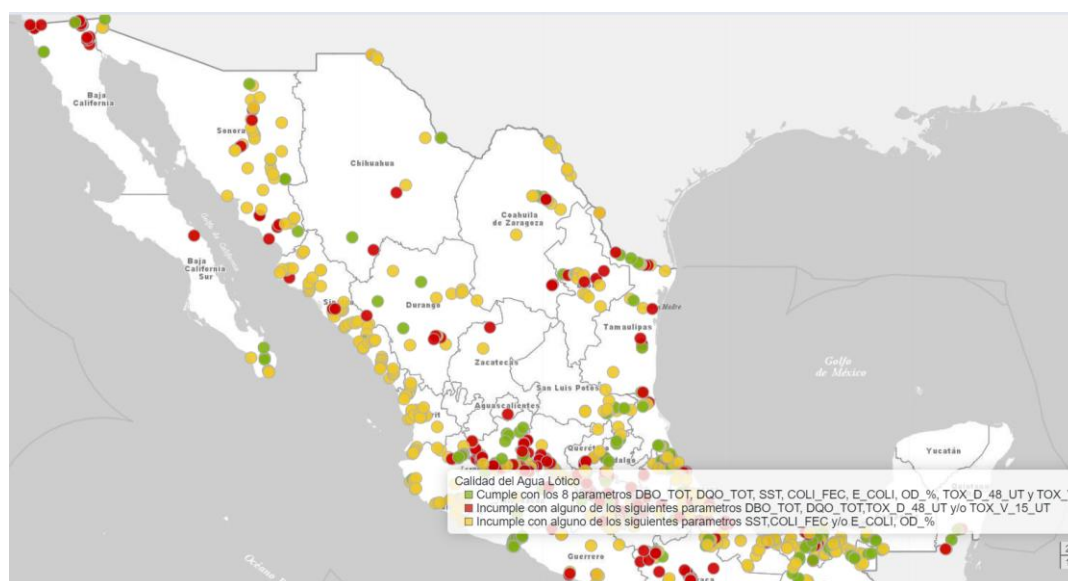


Figura 1-1 Despliegue de los puntos de monitoreo en cuerpos de agua superficiales en México (CONAGUA, 2022).

No obstante, el trabajo que incluye la recolección de muestras, su transporte, análisis y presentación de resultados necesita de numerosos recursos financieros y humanos. Además de los recursos necesarios para obtener los datos, la gestión que implica todos los procesos humanos y analíticos para mantener las bases de datos puede comprometerse. Ante esto, todas las herramientas tecnológicas que

simplifiquen la adquisición o interpretación de los datos producidos, resulta sumamente relevante.

En la literatura técnica, la Inteligencia Artificial ha cobrado importancia por la rapidez y certidumbre con que accede a los resultados, además de los bajos insumos requeridos. Por ello, se han desarrollado campos de aplicación tales como estimación, modelación y simulación de variables físicas, químicas y biológicas. También, se ha utilizado para la reconstrucción o interpolación no asistida de información cuantitativa ausente. Esto es particularmente útil en situaciones donde no se pueden llevar a cabo actividades de muestreo, evitando así la posible disminución en la precisión de los resultados obtenidos.

En zonas muy afectadas por actividades humanas, la IA ofrece una opción para detectar de manera temprana alteraciones en las propiedades o características del recurso hídrico mediante estimaciones de variables clave o de mayor importancia, identificación de las fuentes puntuales e incluso las difusas, así como para la planeación y configuración de estrategias para reducir las afectaciones ambientales.

La Inteligencia Artificial se enfoca en desarrollar modelos que pueden exhibir comportamientos inteligentes. Por lo tanto, los sistemas que utilizan IA son capaces de procesar grandes volúmenes de datos, reconocer patrones y tendencias, y generar predicciones de manera automatizada, rápida y precisa.

Se utilizan modelos estadísticos como árboles de decisión, máquinas de soporte vectorial, gradient boosting y perceptrones multicapa (MLP) para pronosticar parámetros relacionados con la calidad del agua. Para aplicar estos modelos, se

requiere de una gran cantidad de información, que actúa como datos de entrada, y aquí entran en juego las bases de datos sobre la calidad del agua. Cada parámetro medido en la calidad del agua se convierte en una característica de entrada que puede ayudar a describir el comportamiento del parámetro de salida que se desea predecir.

En el presente caso de estudio se calcularán las concentraciones de nitrógeno y fósforo presentes en el área de estudio, el río presidio (véase figura 1-2) haciendo uso del algoritmo para compararlo con la normatividad establecida. A su vez, el suplir con esta deficiencia de datos conllevaría que los muestreos se realicen con menor frecuencia y se disminuiría el gasto económico que provoca su realización en zonas marginadas.



Figura 1-2 Ubicación geográfica de Río presidio (CONSELVA, 2023).

Conocer las concentraciones de nutrientes como el fósforo y nitrógeno que se encuentran presentes en la cuenca hidrológica seleccionada permite poder identificar los posibles impactos ambientales que conllevan, como lo es la eutrofización del

cuerpo de agua y sus debidas consecuencias (aumento de mortalidad en especies de flora y fauna, aumento de la zona anóxica, niveles de toxicidad elevados, entre otros). Además, se busca brindar una alternativa a la deficiencia de datos derivado de problemas para realizar muestreos, lo cual genera dificultades al momento de realizar investigaciones, para esto se utilizará un algoritmo de inteligencia artificial que logre ofrecer estimaciones coherentes utilizando la información obtenida de CONAGUA como datos de entrenamiento.

1.7 Alcance

El proyecto se estima en una duración de 4 meses, teniendo un total de 500 horas en las cuales se entrenará un algoritmo para verificar si este nos puede proporcionar una estimación de la concentración de nitrógeno y fósforo presente en la cuenca hidrológica del río presidio en Sinaloa, México. Este río cuenta con la particularidad de ser represado por la construcción de la presa Picachos.

Los datos que se utilizarán fueron obtenidos de Comisión Nacional del Agua (CONAGUA) con los parámetros de calidad del agua de la cuenca hidrológica “Rio Presidio” ubicando su cuenca alta en el estado de durango para bajar por el estado de Sinaloa, los datos recopilados se comprenden desde diciembre del 2012 hasta noviembre del 2021, teniendo una totalidad de 311 registros de muestreos.

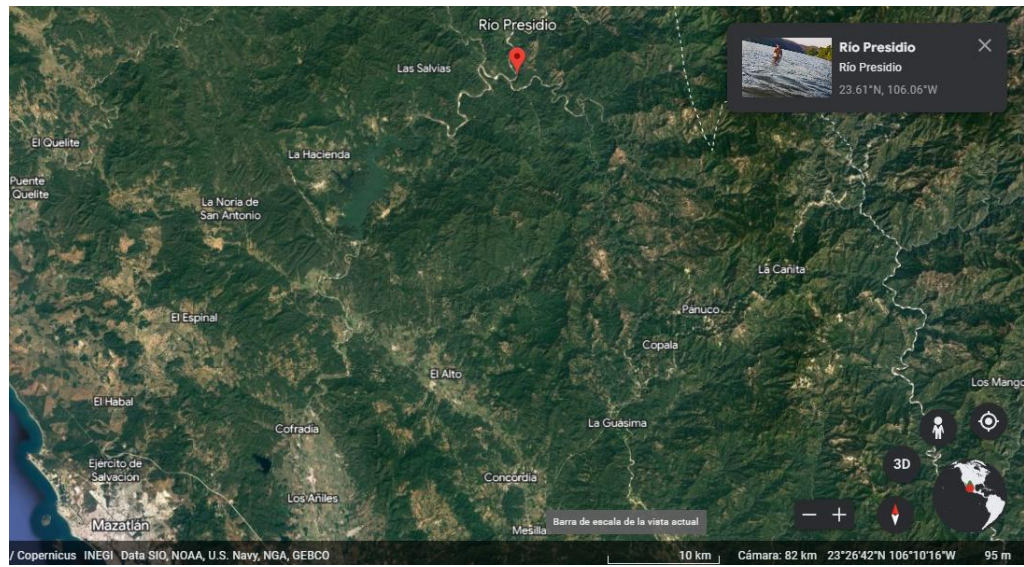


Figura 1-3 Vista satelital de Río Presidio (Google Earth, 2022).

2 FUNDAMENTO TEÓRICO

2.1 Ríos y su temporalidad

2.1.1 Importancia social y económica de los ríos

Los ríos y arroyos se definen como un medio con flujo bifásico ya que transportan agua y sedimento en el mismo flujo (Vide, 2006). Según información de la Comisión Nacional del Agua (Conagua), México posee una red hidrográfica que se extiende por aproximadamente 633 mil kilómetros. Dentro de esta red, destacan 51 ríos que concentran el 87% del escurrimiento superficial y cuyas cuencas abarcan cerca del 65% del territorio nacional. La salud de dichos cuerpos de agua depende de sus características físicas, químicas o biológicas, como el olor, sabor, color, turbidez, conductividad, pH, amonio, presencia de bacterias coliformes, entre otros parámetros que se pueden utilizar para conocer la calidad del agua y en consecuencia determinar la salud del cuerpo de agua.

A lo largo de la historia, el agua ha sido un elemento clave en el surgimiento y consolidación de las civilizaciones. La capacidad de almacenar y transportar agua constituyó la base para su crecimiento, desarrollo y prosperidad. Además de ser esencial para la vida, el agua ha favorecido el progreso económico y comercial de las distintas sociedades. No obstante, el aumento de la población y la industrialización provocaron un incremento en la demanda de agua, así como la expansión de áreas urbanas en regiones con escasez del recurso, lo que generó la necesidad de desarrollar infraestructura para garantizar el suministro a la población. (Monforte García & Cantú Martínez, 2009). Aunque en un principio, las civilizaciones tendieron a asentarse cerca de cuerpos de agua, como ríos, lagos y mares, por instinto de

supervivencia, terminaron beneficiándose directa o indirectamente de los servicios ambientales que brindan. Dichos servicios son procesos ecológicos que se llevan a cabo de manera natural para regular y mantener en funcionamiento a los ecosistemas.

El agua dulce para el consumo humano es uno de los servicios ambientales más importantes y vitales que los ecosistemas acuáticos continentales proveen, pero existen otros no menos importantes (Daily, y otros, 1997). Como mencionó Monforte García (2009) el almacenamiento y la transportación fueron de gran importancia para el desarrollo social y económico y es uno de los servicios contemplados, específicamente como servicio de provisión, también se tienen los servicios de regulación donde se coloca la regulación atmosférica, protección ante la erosión de los suelos, balance hidrológico donde podemos destacar el uso del recurso para actividades antropogénicas como la agricultura, en la misma línea se tienen a los servicios de soporte donde se destaca el sustento para polinizadores, biodiversidad de especies, apoyo a la formación del suelo, reciclaje de nutrientes, defensa natural contra inundaciones, entre otros. Asimismo, se tienen los servicios culturales y turismo donde se comprenden la apreciación estética, la preservación de la identidad cultural, el arraigo a la tierra natal y la vivencia espiritual ligada al entorno natural.

Aunque los ríos benefician directamente a millones de personas, a menudo su importancia permanece en un segundo plano hasta que las alertas rojas aparecen y no pueden cumplir con su propósito, el cual fue inicialmente establecido por el humano y es donde comienzan a producirse crisis económicas, sociales y ambientales. Muchas veces estos propósitos son implantados tomando en cuenta la ubicación geográfica del río o fuente de agua, ya que esto puede afectar a la composición de la misma,

características como contener mayor concentración de cierto componente o la carencia de alguno puede deberse a que en regiones más húmedas suele tener más nutrientes y sedimentos mientras que el agua llevada por ríos en zonas montañosas o altas tienen tendencia a fluir con mas velocidad provocando una mayor erosión a sus alrededores y por consecuencia llevar mas sedimentos y minerales en su contenido. Dependiendo las características del agua es el uso que el humano le dará o en su defecto se le proporcionará un tratamiento para hacerla apta para su utilización en actividades económicas o su consumo.

2.1.2 Época de lluvia y estiaje en ríos

El clima del estado de Sinaloa es en un 48% cálido subhúmedo y se encuentra al noroeste-sureste desde el municipio de Choix hasta los límites con Nayarit, un 40% es seco y semiseco presente desde El Fuerte hasta Mazatlán, un 10% es muy seco y se encuentra en Los Mochis mientras que un 2% ubicado en las partes altas de la Sierra Madre Occidental es templado tal como se muestra en la figura 2-1.



Seco y semiseco	40%*
Cálido subhúmedo	48%*
Muy seco	10%*
Templado subhúmedo	2%*

Figura 2-1 Carta de climas en Sinaloa usando la clasificación propuesta por la SEP. (INEGI, 2023).

Las lluvias en el estado comienzan a finales del mes de junio y concluye a principios de octubre según los registros obtenidos en la Estación Climatológica de la Facultad de Biología de la Universidad Autónoma de Sinaloa (UAS). Durante la época de lluvia se abastecen las presas, ríos, arroyos y acuíferos del territorio, lo cual proporciona de agua suficiente para sobrellevar la época de estiaje después de la etapa de agricultura de secano (Centro Nacional de Prevención de Desastres , 2018). En ocasiones la cantidad de agua que aporta la lluvia sobrepasa la capacidad máxima de los cauces provocando que los ríos se desborden generando una acumulación de agua en zonas que generalmente están secas a lo cual se le conoce como inundación.

Por otro lado, aguas arriba en la zona de estudio (río presidio) se tiene al estado de durango donde un 40% presenta clima seco y semiseco, 34% clima templado subhúmedo, un 14% clima muy seco, el 11% cálido subhúmedo y en el restante 1% templado húmedo. Las lluvias se presentan en verano, principalmente en los meses de julio y agosto donde la precipitación media del estado es de 500 mm anuales.

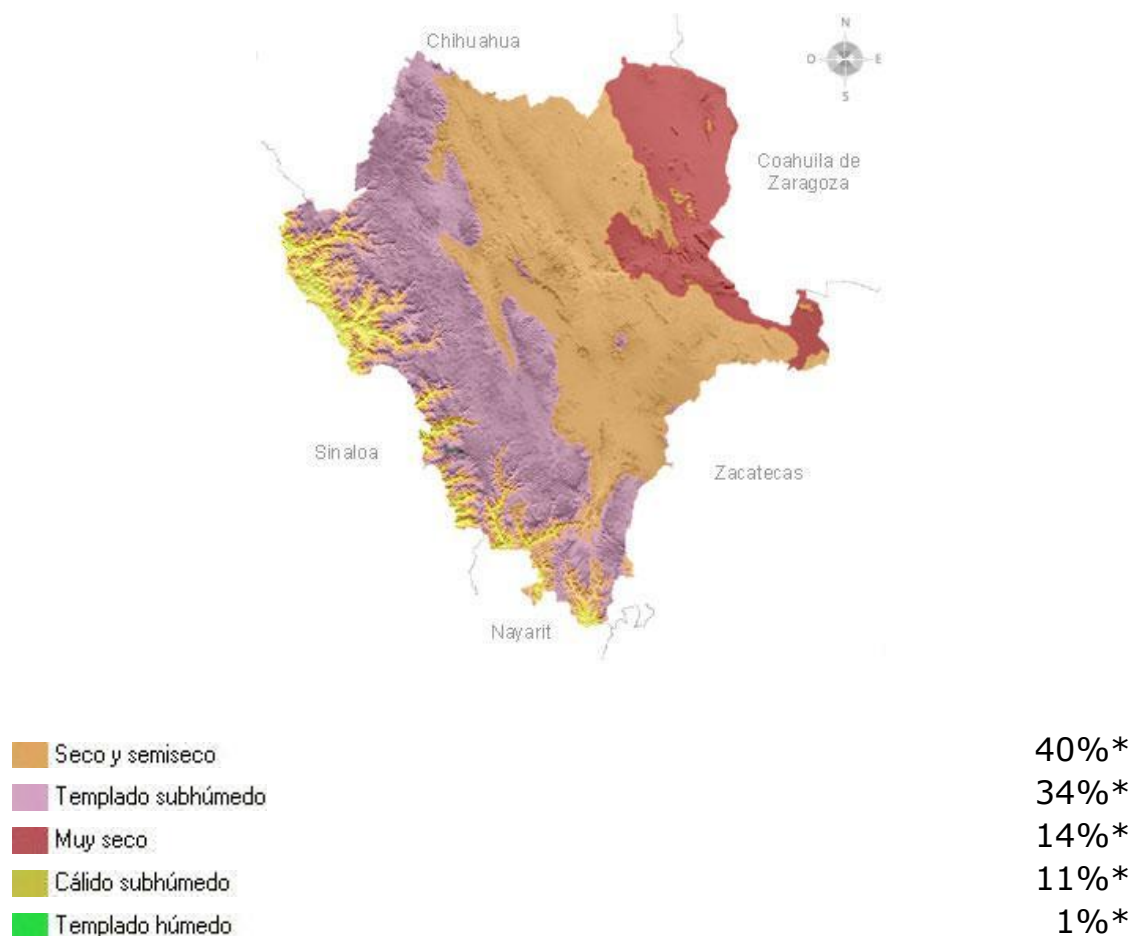


Figura 2-2 Carta de climas del estado de Durango 1:1 000 000 (INEGI, 2019).

El estiaje se define como el nivel de caudal mínimo que alcanzan los ríos, lagunas o acuíferos y se da durante la época de mayor calor como consecuencia de la época de sequía, causada por falta de lluvias derivado al calentamiento global o por una inadecuada administración del agua e incluso por un deficiente manejo de las presas.

En otras palabras, cuando se menciona que un río atraviesa un periodo de estiaje, se hace referencia a que mantiene un caudal reducido durante una época seca prolongada. La medición precisa de estos caudales es fundamental para el desarrollo de proyectos orientados al manejo de los corredores fluviales, como los estudios de restauración y re-naturalización de los cauces. (Mendiondo, Clarke, R., & Tönsmann, F, 2000), determinación de caudales ecológicos (Doyle, Stanley, E., Strayer, D., Jacobson, R, & Schmidt, J., 2005), requerimientos de hábitat, formulación de balances hídricos y al análisis y mitigación de sequías (García, Vicario, Colladon, Ravelo, & Rodríguez).

También, se debe a que en esta época del año hay una mayor insolación que provoca la evapotranspiración de las plantas y la evaporación más intensa de los cuerpos de agua. El estiaje es un problema grave particularmente en el estado de Sinaloa debido a que es un estado agrícola y tiene efectos negativos sobre la producción de cultivos durante esta época.

2.1.3 Impacto por construcción de presas

En la actualidad, son pocos los ríos que conservan su estado natural, libre y sinuoso, debido a la creciente demanda de energía hidroeléctrica, riego y navegación interior, lo que ha impulsado la expansión de presas e infraestructura fluvial, fragmentando y alterando los cursos de agua (UNEP, 2022). Un estudio realizado en 2019 por el Fondo Mundial para la Naturaleza (WWF) reveló que casi dos tercios de los ríos más largos del mundo han sido obstruidos por presas. Esta intervención tiene efectos significativos sobre la calidad del agua, que se ven reflejados desde el embalse hasta la desembocadura del río. Entre los parámetros más afectados se encuentran la

reducción del oxígeno disuelto, cambios en la temperatura, así como la estratificación y acumulación de sedimentos (Hispagua, Sistema Español de Información sobre el Agua, 2007).

2.1.4 El recurso agua en México

2.1.4.1 Precipitación

Según la ley de conservación de la materia de Lavoisier, la cantidad total de agua en la Tierra no ha cambiado, pero la escasez de agua ha aumentado en la mayor parte del mundo. Hay una razón real para esta suposición, y los cambios en el ciclo del agua y los problemas de contaminación han reducido la cantidad de agua dulce en algunas áreas. (Monforte García & Cantú Martínez, 2009). Esta situación se refleja en el caso de México, donde en los últimos años han sido frecuentes los reportes sobre escasez de agua. Además, los datos de precipitación anual muestran esta tendencia: durante el periodo 1981-2010, la precipitación normal del país fue de 740 mm, mientras que en 2019 y 2020, el mes de abril registró los valores más bajos, con 6.6 mm y 14 mm, respectivamente. La precipitación acumulada en todo el territorio mexicano del 1° de enero al 31 de diciembre de 2020 alcanzó 722.5 mm, lo que representa un 2.4% menos que la media histórica del periodo 1981-2010 (740 mm) (CONAGUA, 2021).

2.1.4.2 Abastecimiento hídrico de las principales actividades económicas en México.

El agua es un recurso esencial para llevar a cabo diversas actividades humanas, ya sea para subsistir o bien, utilizarla en producir bienes y servicios dentro de las actividades económicas, sin embargo, actualmente el recurso se ve amenazado por la sobreexplotación.

En México el crecimiento poblacional ha ido de la mano con un crecimiento económico y aunque puede verse como algo positivo es crucial prestar atención a las alertas sobre la insuficiencia de agua que han surgido como resultado. La densidad de la población y el gasto de agua en actividades económica han provocado escasez, llegando a afectar zonas donde el recurso hídrico no se visualizaba como un problema debido a sus reservas y fuentes disponibles.

En la actualidad, México destaca a nivel mundial por la exportación de vegetales y frutos, al punto que la economía nacional ha disminuido su dependencia del petróleo y se sustenta de manera significativa en la venta de productos agrícolas y en el turismo. En otras palabras, el país ha dejado de exportar principalmente petróleo y ahora transfiere de manera indirecta agua y energía a través de sus productos agrícolas, industriales y de servicios. Sin embargo, esta dinámica aumenta la vulnerabilidad de las regiones donde el agua subterránea es la principal fuente para la producción de bienes de autoconsumo y para la exportación, frecuentemente superando el potencial sostenible de estos recursos. (Díaz-Delgado & Moeller-Chavez, 2019).

Se estima que cerca del 76 % del recurso en México está destinado a la agricultura mientras que el 14 % al abastecimiento público, el 5 %, en las termoeléctricas y el 5 % restante en la industria (INEGI, 2024). Sin embargo, se puede observar esta diferencia de manera mas clara en la Figura 2-3.

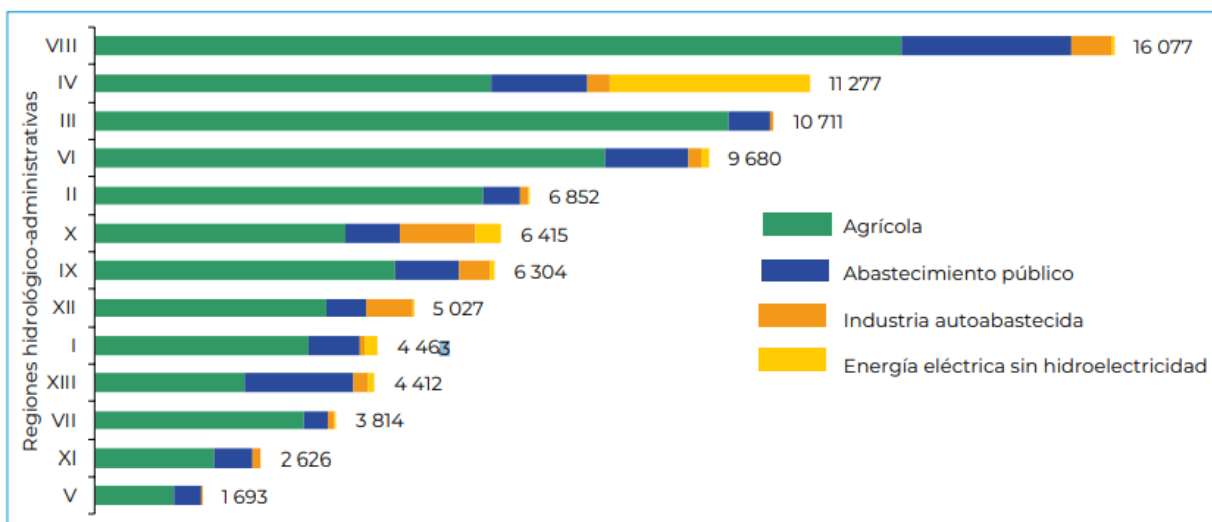


Figura 2-3 Volumen concesionado para usos consuntivos, 2019 (hm³) (CONAGUA, 2021).

2.2 Importancia de los nutrientes

2.2.1 Problemática del Nitrógeno en cuerpos de agua

El nitrógeno es el séptimo elemento en la tabla periódica, conforma parte de la atmosfera terrestre en un 78% en forma de nitrógeno gaseoso y es uno de los elementos más comunes debido a su abundancia en su forma molecular, además, en el ámbito ambiental es conocido como uno de los nutrientes minerales más importantes debido a su importante desempeño para el crecimiento adecuado de las plantas y, por ende, cumple un papel crucial en la productividad primaria de la mayoría de los ecosistemas.

Se denomina sustancias o compuestos nitrogenados a las biomoléculas que contienen nitrógeno, ya sea macromoléculas o productos de desecho. Las macromoléculas nitrogenadas con mayor importancia biológica son los ácidos nucleicos y las proteínas; sus precursores son las bases nitrogenadas y los aminoácidos.

En el agua natural, además del nitrógeno molecular disuelto, pueden encontrarse diversos compuestos nitrogenados, tales como amoníaco, nitritos, nitratos y distintas moléculas orgánicas que contienen nitrógeno, entre ellas los aminoácidos, la urea, la hidroxilamina, las amidas y ciertos derivados de la piridina, entre otros. (Ambientum, 2022).

Los compuestos nitrogenados adquieren importancia no solo porque el nitrógeno constituye un elemento esencial en las funciones metabólicas de los organismos vivos, sino también por su amplia presencia en el aire, el suelo y el agua. Esto se debe a la compleja química del nitrógeno, capaz de adoptar siete estados de oxidación diferentes (de -3 a +5), lo que le permite formar una gran variedad de compuestos. Entre los más comunes se encuentran el amoníaco (NH_3) y el ion amonio (NH_4^+) con estado -3, el nitrógeno gaseoso (N_2) con estado 0, el óxido de nitrógeno (N_2O_3) con estado +3 y el pentaóxido de dinitrógeno (N_2O_5) o ácido oxonítrico (HNO_3) con estado +5. La amplia diversidad de compuestos nitrogenados se explica por las diferentes etapas del ciclo del nitrógeno, a través de las cuales este elemento experimenta transformaciones que regulan su presencia en múltiples formas químicas. Sin embargo, dicho ciclo ha sido alterado por actividades humanas como fertilizar excesivamente los suelos, agricultura intensiva, tala de árboles, uso excesivo de combustibles fósiles entre otros, que han aumentado casi al doble la concentración de N fijado anualmente en la biósfera. El incremento de la fijación de este elemento ha logrado alcanzar niveles tóxicos en agua lo cual altera la capacidad de supervivencia, crecimiento y reproducción de los animales, el nitrógeno llega a los diferentes cuerpos de agua en la forma de amonio, nitrato y nitrito, aunque el más tóxico resulta ser el

amonio no ionizado (NH_4) y es a la vez el más común debido a su origen de descomposición de materiales proteicos (Cárdenas Calvachi & Sánchez Ortiz, 2007).

2.2.2 Problemática del Fósforo en cuerpos de agua

Es un elemento no metálico perteneciente a la familia del nitrógeno (grupo Va de la tabla periódica). El fósforo es un elemento indispensable para el crecimiento y desarrollo de las plantas, su rol no puede ser realizado por ningún otro nutriente, desempeña tareas en los procesos de transporte de energía en la planta donde los fosfatos de alta energía (adenosina difosfato (ADF) y adenosina trifosfato (ATF)) son la fuente de energía que empuja una multitud de reacciones químicas dentro de la planta. El fósforo es absorbido por la planta principalmente como ion ortofosfato primario (H_2PO_4^-), pero también se absorbe como ion fosfato secundario (HPO_4^-), la absorción de esta última forma se incrementa a medida que se sube el pH (Sultenfuss & Doyle, 1999).

La contaminación debido a este elemento en el agua se debe principalmente a su uso excesivo en fertilizantes agrícolas que posteriormente llegan a los cuerpos de agua debido a la erosión de los suelos y a la descarga de materia orgánica derivadas de urbes y granjas de animales domésticos. La fijación del fósforo con el suelo es bastante efectiva, siendo así una de las razones centrales de su abuso en actividades agrícolas, pero incluso con su gran capacidad de fijación una cantidad, aunque sea mínima logra llegar a los cuerpos de agua y estas pequeñas proporciones son suficientes para tener severos impactos en la calidad de las fuentes de agua superficiales (SEMARNAT, 2023).

2.2.3 Impactos ambientales de nutrientes en exceso

En los ecosistemas acuáticos, los nutrientes constituyen elementos fundamentales para el metabolismo de los organismos vivos, ya que forman parte de su estructura y funcionamiento. Estos compuestos son esenciales en el inicio de la cadena trófica, pues favorecen el crecimiento del fitoplancton y, en consecuencia, el desarrollo de peces y crustáceos. No obstante, cuando su concentración excede los niveles naturales, pueden generar desequilibrios y provocar procesos de contaminación en el medio acuático. (Cárdenas Calvachi & Sánchez Ortiz, 2007).

El nitrógeno (N) y el fósforo (P) son los nutrientes que con mayor frecuencia limitan el crecimiento de las plantas si su disponibilidad no es adecuada, por el contrario, al encontrarlos en exceso estos favorecen la reproducción y crecimiento desmedido de algunos organismos vegetales (Faustino, Graciano, Gortari, & Guiamet, 2011). Esto fomenta el desarrollo de un fenómeno denominado como eutrofización, palabra que viene del griego “eutros” la cual significa "bien nutrido".

La eutrofización puede entenderse como un proceso de deterioro de la calidad del agua ocasionado por el enriquecimiento excesivo de nutrientes en cuerpos de agua superficiales, como ríos, lagos y embalses. Un ambiente eutrofizado se caracteriza por la presencia de niveles anormalmente altos de nutrientes, siendo el fósforo y el nitrógeno los principales responsables de este fenómeno (Ministerio para la Transición Ecológica y el Reto Demográfico, s.f.). Esta condición favorece un incremento en la densidad del fitoplancton, un conjunto de organismos planctónicos que incluye algas microscópicas como diatomeas, dinoflagelados y cianobacterias. El aumento de fitoplancton altera el comportamiento del ecosistema acuático, generando pérdida de

calidad del agua y condiciones de anoxia. Si bien se trata de un proceso natural que ocurre a lo largo de la vida de los cuerpos de agua, normalmente tardando miles de años, en la actualidad se ha acelerado considerablemente debido a las actividades humanas. (Lozano, 2016).

Cuando ocurre de manera natural, la eutrofización es un proceso progresivo y de desarrollo lento. En cambio, la eutrofización artificial o cultural se manifiesta de forma acelerada, generando un incremento descontrolado en la producción de biomasa fitoplanctónica. Este aumento supera la capacidad del sistema acuático para asimilar dicha materia orgánica al mismo ritmo, lo que ocasiona un marcado desequilibrio ecológico. (Bem, Borba Braga, & Rodrigues de Azevedo, 2013).

Para evitar la eutrofización es necesario mantener un monitoreo constante en la calidad del agua, principalmente en las concentraciones de nitrógeno y fósforo, además de estimar la concentración de biomasa contenida en el cuerpo de agua, la cual se encontrará en forma de fitoplancton (Bem, Borba Braga, & Rodrigues de Azevedo, 2013).

2.2.4 Monitoreo De Calidad De Agua

Hoy en día, hay una creciente preocupación por el deterioro del medioambiente y el impacto que ciertas actividades humanas tienen sobre él. En particular, se destacan los factores que afectan los recursos naturales, especialmente el agua, un elemento fundamental para la vida. La contaminación de lagos y ríos sigue siendo un problema sin resolver, ya que los ecosistemas acuáticos han acumulado numerosos residuos provenientes de actividades humanas a lo largo del tiempo. Por lo tanto, es

esencial monitorear la calidad del agua y detectar los vertidos de efluentes contaminantes para poder hacer un seguimiento histórico de su evolución, tomar medidas preventivas y desarrollar soluciones que, en la medida de lo posible, mitiguen el impacto negativo de las actividades humanas en estos ecosistemas acuáticos (Lupi, Canziani, & Zaradnik, 2020).

2.2.5 Red De Monitoreo De Conagua

Desde el año 2000 la Comisión Nacional del Agua con apoyo de la Red Nacional de Monitoreo de la Calidad del Agua (RNMCA) gestiona una red de vigilancia de la calidad del agua que incluye 5,000 puntos de monitoreo repartidos en seis categorías: aguas superficiales, aguas subterráneas, estudios especiales, zonas costeras, descargas superficiales y descargas subterráneas. De estos, la red de aguas superficiales, que se centra en el análisis del estado de ríos, lagos y embalses, comprende 2,514 puntos de monitoreo (Díaz-Delgado & Moeller-Chavez, 2019). Los monitoreos toman en cuenta 46 parámetros de calidad del agua entre los que destacan los relacionados con materia orgánica disuelta, nutrientes, microorganismos patógenos, parámetros físicos y metales pesados, entre otros.

2.3 Inteligencia Artificial

2.3.1 Inteligencia artificial y la predicción de datos

El machine learning es una rama de la inteligencia artificial que tiene como objetivo principal capacitar a los algoritmos para aprender y extraer conocimiento a partir de datos. Se divide en dos categorías principales: aprendizaje supervisado y no supervisado. En el aprendizaje supervisado, se lleva a cabo un proceso de entrenamiento que implica la observación de las características clave de un evento,

que actúan como datos de entrada. Luego, en la fase de prueba, estas características se comparan con nuevos datos, dando como resultado un modelo que ha sido entrenado y evaluado para explicar los datos de entrada (Riabani Mercado, García Fernández, & Herrera Acebey, 2016). Es importante recalcar que una vez que se ha entrenado un modelo de aprendizaje automático, se puede utilizar para hacer predicciones sobre nuevos datos.

Para realizar las predicciones existen una gran cantidad de técnicas. En general, para abordar cualquier problema que se presente, no hay una única técnica para resolverlo; en cambio, es posible adoptar diferentes enfoques para su solución. A continuación, se mencionan brevemente algunas de las opciones de técnicas con las que se cuentan actualmente (Aluja, 2001).

2.3.2 Aplicaciones de la ciencia de datos en calidad del agua

José Carbajal y Luis Sánchez, en su estudio “Modelo basado en redes neuronales artificiales para la evaluación de la calidad del agua en sistemas de cultivo extensivo de camarón”, desarrollaron un modelo computacional novedoso para evaluar la calidad del agua en granjas de cultivo extensivo de *Litopenaeus vannamei*. El monitoreo se llevó a cabo por el Instituto Tecnológico de Sonora (ITSON) en colaboración con el Centro de Investigaciones Biológicas del Noreste (CIBNOR), en la granja camaronera “Gez Acuícola”, ubicada en el ejido “Rancho Chapo”, Huatabampo, Sonora, México.

Para el estudio se seleccionaron cuatro parámetros: temperatura, oxígeno disuelto y salinidad. Mediante el uso de una red neuronal artificial (RNA), se asigna una respuesta a cada parámetro, lo que permite relacionar un conjunto de concentraciones con un estado de calidad del agua, generando un indicador global clasificado como excelente, bueno, regular o deficiente. Además, la RNA procesa y aprende las relaciones y dependencias intrínsecas entre los parámetros, lo que mejora la precisión de las evaluaciones realizadas. Se concluyó que las redes neuronales artificiales son una buena opción para la detección de estados óptimos o no deseados para un buen manejo del agua en este tipo de cultivos.

Siguiendo el ejemplo de la aplicación de redes neuronales, se cuenta con el ejemplo de Samira Nemati en “Artificial Neural Network Modeling of Total Dissolved Solid in the Simineh River, Iran” quien en 2014 explicó como logró modelar las concentraciones de sólidos disueltos totales (TDS) en el río Simineh al norte de Irán con el apoyo de una red neuronal artificial y así evaluar la calidad del agua presente en dicho cuerpo, los parámetros utilizados fueron cantidades de cloro, magnesio, calcio, sodio, bicarbonato, sulfato y caudal con una base de datos alimentada desde 1993 hasta 2011, también se hace uso de la ecuación de Garson para reconocer la importancia de las diferentes combinaciones de las variables y hacer uso de la combinación indicada, los resultados indicaron una predicción favorable y precisa obtenida por el modelo red neuronal artificial utilizado.

2.3.3 Modelos predictores

En la actualidad, existe una amplia variedad de modelos que permiten realizar predicciones con distintos niveles de precisión. Según la revisión bibliográfica

consultada, las diversas técnicas de machine learning descritas en la literatura corresponden a evoluciones o adaptaciones de métodos previamente conocidos. El aprendizaje automático (machine learning), comúnmente abreviado como ML, es una rama de la inteligencia artificial (IA) que tiene la capacidad de “aprender” y mejorar su desempeño con el tiempo. A diferencia de los sistemas tradicionales basados en reglas fijas, esta tecnología reconoce patrones en los datos de entrada y emplea algoritmos que se ajustan y evolucionan progresivamente. (Aruna Flarence, 2018).

El concepto de análisis predictivo se refiere al uso de una o varias técnicas estadísticas, o de aprendizaje automático, con el fin de generar una predicción cuantitativa acerca de lo que ocurrirá en el futuro, para lograr un análisis predictivo con inteligencia artificial se utilizan los denominados modelos predictivos. Para llevar a cabo el análisis predictivo es indispensable disponer de una considerable cantidad de datos, tanto actuales como pasados, para poder establecer patrones de comportamiento y así inducir conocimiento (Timón, 2017)

Existen 2 grandes tipos de modelos predictivos; los de clasificación y los de regresión, donde, de manera resumida los de clasificación permiten predecir la pertenencia a una clase y los de regresión permiten predecir valores.

Los modelos de clasificación suelen partir al conjunto de datos en dos conjuntos de datos más pequeños: entrenamiento y prueba. El subconjunto de datos de entrenamiento es utilizado para estimar los parámetros del modelo y el subconjunto de datos de prueba se emplea para comprobar el comportamiento del modelo estimado (Martín-Romero, 2020).

Los más comunes se enlistan a continuación:

- Análisis Discriminante

El análisis discriminante, propuesto por Fisher en 1936, es una técnica estadística empleada para predecir la pertenencia de un elemento a un grupo determinado por una variable dependiente, utilizando para ello un conjunto de variables independientes. Su objetivo principal es identificar las diferencias existentes entre los grupos y estimar la categoría o clase a la que pertenece un objeto, en función de los valores que presentan dichas variables independientes. (Casanova, 2022).

- Clasificador Bayesiano

El aprendizaje bayesiano se basa en identificar la hipótesis más probable mediante el uso de un conjunto de datos de entrenamiento y la estimación previa de la probabilidad de cada hipótesis. Los modelos de clasificación supervisada que se fundamentan en este enfoque reciben distintos nombres, entre ellos idiot Bayes, naive Bayes, simple Bayes y Bayes independiente. El enfoque naive Bayes surgió en la década de 1980, y su funcionamiento puede explicarse a partir del teorema de Bayes.

Consideremos dos sucesos aleatorios, A y B, cuyas probabilidades se denotan como $p(A)$ y $p(B)$, con $p(B) > 0$. Si se conocen de manera previa las probabilidades de A y B, así como la probabilidad condicional de B dado A, entonces la probabilidad de A dado B, $P(A|B)$, puede calcularse utilizando la fórmula correspondiente. Este teorema requiere contar con información a priori; de no estar disponible, estas probabilidades

deben estimarse, lo cual incrementa el costo computacional (Santos, Dallos, & Gaona-García, 2020).

Por otro lado, los modelos de regresión se utilizan para analizar y cuantificar la relación entre una variable dependiente continua y una o más variables independientes. Su objetivo es determinar en qué medida la variable dependiente puede explicarse por los predictores y, además, permitir la predicción de nuevos valores de la variable dependiente a partir de las variables independientes.es (Martín-Romero, 2020).

Algunos modelos de regresión serían:

- Regresión lineal

Un modelo de regresión lineal permite analizar y cuantificar la relación entre una variable dependiente (U) y una o varias variables independientes (C) mediante la formulación de una ecuación de tipo lineal. (Martín-Romero, 2020).

$$Y = \beta_0 + \sum \beta_i X_i$$

Donde β representa las estimaciones de parámetros lineales que se ajustan para que la medida sea óptima.

- Regresión no lineal

Un modelo de regresión no lineal permite construir una función que describa las relaciones no lineales entre una variable dependiente (U) y una o varias variables independientes (C), generando así una ecuación de carácter no lineal. A diferencia de la regresión lineal, este tipo de función no depende de parámetros lineales específicos,

lo que le permite adoptar múltiples formas según la naturaleza de los datos y la relación entre las variables. (Martín-Romero, 2020).

$$Y = f(X, \beta)$$

Donde f es alguna función no lineal respecto unos parámetros β no lineales. Las formas no lineales posibles incluyen: convexa, cóncava, curva sigmoideal, curvas asintóticas, etc.

- Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNAs) constituyen un enfoque para la resolución de problemas, ya sea de manera independiente o combinadas con otros métodos, especialmente en tareas de clasificación o regresión donde la relación entre datos y conocimiento se inclina hacia los datos. Este enfoque resulta útil cuando se requiere aprendizaje en tiempo real y cierta tolerancia a fallos. Las RNAs se adaptan de forma dinámica, ajustando continuamente los “pesos” de sus interconexiones para mejorar su desempeño. Su funcionamiento se inspira en el comportamiento del cerebro humano y del sistema nervioso, donde redes de neuronas biológicas con capacidades individuales limitadas logran desarrollar habilidades cognitivas complejas gracias a la extensa conectividad entre ellas. (Salas, 2004).

- Árboles de decisión

Los árboles de decisión o de clasificación son un modelo de Machine Learning que, a partir de una base de datos, genera diagramas de construcciones lógicas que facilitan la resolución de problemas. Estos se componen de nodos, ramas y hojas. Los nodos

son las variables de entrada, las ramas representan los posibles valores de las variables de entrada y las hojas son los posibles valores de la variable de salida (Parra, 2019).

- Random Forest

Este procedimiento de aprendizaje automático emplea bosques de decisión aleatorios compuestos por un grupo de árboles de decisión, dichos bosques se generan a través de un algoritmo que incluye una dosis de azar para disminuir la relación entre los árboles. Este algoritmo tiene un historial muy positivo, ya que se clasifica como una técnica de aprendizaje supervisado. Su propósito es crear múltiples árboles de decisión basados en un conjunto de datos de entrenamiento, que producirá predicciones según lo que se desee ajustar (Patiño Pérez, Silva Bustillos, Munive Mora, & Botto Tobar, 2020).

2.3.4 Métricas de evaluación

2.3.4.1 Coeficiente de determinación

El coeficiente de determinación es una medida estadística que examina qué tan bien la variación en una variable se explica por la variación en otra variable al predecir el resultado de un evento. En otras palabras, este coeficiente, conocido como R cuadrado (R^2) evalúa la fuerza de la relación lineal entre dos variables y es ampliamente utilizado por los investigadores cuando trabajan en análisis de tendencias.

La correlación, con este tipo de datos, se calcula simplemente con elevar al cuadrado el coeficiente de correlación de Pearson o Spearman

$$R^2 = \frac{\sigma^2_{XY}}{\sigma^2_X \sigma^2_Y}$$

Donde:

- σ_{XY} es la covarianza de (X,Y)(X,Y).
- σ^2_X es la Varianza de la variable X.
- σ^2_Y es la Varianza de la variable Y.

2.3.4.2 Raíz del Error Cuadrático Medio (RMSE)

RMSE, que significa *Root Mean Squared Error* (Raíz del Error Cuadrático Medio), es una métrica utilizada para cuantificar la magnitud del error entre dos conjuntos de datos. Su principal ventaja es que expresa el error en la misma unidad que la variable dependiente, lo que facilita la interpretación de los resultados. En términos prácticos, permite comparar valores predichos con valores observados o conocidos. Esta métrica, también conocida como Raíz de la Desviación Cuadrática Media, es ampliamente empleada en sistemas de información geográfica (SIG) y en diversas aplicaciones de modelado y predicción.

Su fórmula es la siguiente: $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$

donde:

- n es el número total de observaciones en el conjunto de datos.
- y_i es el valor real de la variable dependiente para la observación i.

- y^i es el valor predicho por el modelo para la observación i .

2.3.4.3 Desviación Absoluta Media

La Desviación Absoluta Media (MAE, por sus siglas en inglés) es una métrica empleada para evaluar la precisión de modelos de regresión. Se obtiene sumando las diferencias absolutas entre los valores observados y los valores predichos, y dividiendo el resultado entre el número total de registros. Esta estadística de error refleja, en promedio, la distancia entre cada par de puntos de datos reales y los valores ajustados por el modelo, proporcionando una medida clara de su exactitud. El MAE proporciona una medida de la magnitud promedio de los errores de predicción en términos de la variable original. A diferencia del MSE y RMSE, el MAE no penaliza de manera exponencial los errores más grandes, lo que lo hace menos sensible a valores atípicos.

3 MATERIALES Y MÉTODO

3.1 Obtención y exploración de base de datos

Se realizó la solicitud de la base de datos a la Comisión Nacional del Agua, la cual, contiene información recopilada desde 2012 hasta 2021 de los 11 puntos de muestreo en donde se puede encontrar la siguiente información: Clave Sitio, Clave Monitoreo, Nombre Del Sitio, Tipo De Cuerpo De Agua, Fecha de muestreo , Coliformes Fecales, Coliformes Totales, E Coli, Carbono orgánico total, Carbono orgánico total contenido en los sólidos en suspensión, DBO contenida en los sólidos en suspensión, DBO total, DQO contenida en los sólidos en suspensión, DQO total, Amoníaco total (N_Nh3), Nitritos (N_No2), Nitratos (N_No3), Nitrógeno orgánico, Nitrógeno total, Nitrógeno Total Kjeldahl, Toxicidad con duración de 48 horas, Toxicidad con duración de 15 minutos, Toxicidad con duración de 30 minutos, Fósforo Total (P_Tot), Fosfato Reactivo o Fosfato Ortofosfórico (Orto_PO4), Color verdadero, Absorbancia UV, Solidos disueltos totales, Conductividad campo, pH, OD (%), OD (Mg/L), Sustancias Activas Aniónicas Móviles (SAAM)., Solidos suspendidos totales, Turbiedad, Arsénico Total, Cadmio Total, Cromo Total, Mercurio Total, Níquel Total, Plomo Total, Cianuro Total, Dureza Total, Temperatura ambiente, Temperatura del agua y Caudal.

También se realizó una exploración de la base de datos, la cual cuenta con 46 características (parámetros) en total y 378 instancias (datos). Dicho conjunto contiene los siguientes puntos de muestreo (con su respectiva clave sitio): Presa Picachos (OCPNO4614), El Walamo (OCPNO4624), Pozo Villa Unión (OCPNO4621), Walamo (OCPNO4619), Rio Presidio 5 (OCPNO4615), Rio Presidio 4 (OCPNO4616), Rio

Presidio 3 (OCPNO4617), Rio Presidio 2 (OCPNO4620), Rio Presidio 1 (OCPNO4622), Puente Villa Unión (OCPNO4623), Pozo CNA 82 (OCPNO4618), Desembocadura Rio Presidio (OCPNO4627). Fueron retirados de la base de datos aquellos puntos de muestreo que contenían agua subterránea (Pozo Villa Unión, Walamo y Pozo CNA 82) ya que tanto agua superficial como agua subterránea cuentan con diferencias significativas en términos de origen, composición química e interacción con el entorno por lo que su presencia en el conjunto de datos podría generar inconsistencias o perturbaciones durante el entrenamiento del algoritmo. Además, el conjunto de datos fue dividido en 2 tomando en cuenta la fecha en la que fue realizado el muestreo con el objetivo de obtener un conjunto de datos para la temporada de estiaje y otro para la temporada de lluvia. Al finalizar la separación se obtuvo un total de 122 instancias para el conjunto de lluvia y 213 para el conjunto de estiaje de los 8 puntos de muestreos seleccionados.

3.2 Preprocesamiento de Datos

El conjunto de datos contenía muchos datos “NaN” o “datos faltantes” por lo que fue necesario realizar una técnica para procesar dicho problema ya que su presencia puede generar ruido durante el entrenamiento y afectar de manera negativa a la predicción. Se le solicitó al algoritmo crear una instancia de “SimpleImputer” (Una clase SimpleImputer de la biblioteca scikit-learn) con la estrategia de imputación configurada en 'mean', lo que significa que se utilizará la media de cada columna para rellenar los valores faltantes. Con el método de fit_transform se calculó la media de cada columna para rellenar los valores faltantes. Una vez aplicado el código este imputa los valores faltantes en las columnas numéricas del conjunto de datos utilizando la media de cada

columna como estrategia de imputación. La imputación es una práctica común para lidiar con valores faltantes en conjuntos de datos antes de aplicar modelos de aprendizaje automático.

Además, se aplicó una normalización de datos para apoyar en el escalamiento de las variables, esto ayuda a que todas tengan rangos comparables ya que, si las variables no están en la misma escala, algunas pueden tener un impacto desproporcionado en el modelo. La normalización de datos fue realizada con la función “sklearn.preprocessing” y su función de “MinMaxScaler” lo que asegura que después de la ejecución todos los valores se encuentren dentro del rango de 0 a 1.

3.3 Selección de Características Predictoras

Este paso se realizó partiendo de la importancia de las características y esta fue distinta para cada conjunto de datos, hasta este punto se estuvieron manejando paralelamente 4 conjuntos de datos, 2 para la estimación de Nitrógeno (uno para la temporada de estiaje y otro para la temporada de lluvia) y 2 para la estimación de Fósforo en la cuenca seleccionada. Este paso fue realizado con la ayuda del modelo de RandomForest donde se calcula la importancia de cada característica en el proceso de toma de decisiones del modelo, esto se realiza con el comando `forest.feature_importances_` disponible en la biblioteca del modelo. La importancia se evalúa según cómo cada característica contribuye a reducir la impureza en los nodos del árbol de decisión, que es la base del funcionamiento de RandomForest. Una vez el modelo calcula la importancia de cada característica utiliza las primeras 20 más importantes para la predicción de la variable objetivo, es de gran utilidad ya que

también toma en cuenta las interacciones complejas que pueda existir entre las características.

Además, se eliminaron aquellas características que tuvieran involucrada de manera directa a la variable objetivo, como por ejemplo para la predicción del Nitrógeno total se removieron las características Nitritos, Nitratos y Nitrógeno Kjeldahl debido a su relación con la misma. Lo mismo con el fósforo como variable objetivo y la eliminación de la variable “Fósforo Ortofosfórico” de sus respectivos conjuntos de datos.

3.4 Entrenamiento de inteligencia artificial

Se utilizó el modelo de regresión de RandomForest para entrenar el sistema. Este modelo es capaz de manejar datos complejos y no lineales, lo que lo hace adecuado para problemas de regresión en el contexto de la calidad del agua. Para entrenarlo se separaron los datos en “datos de entrenamiento” y “datos de prueba” los cuales están conformados por el 80-90% y el 10-20% de los datos respectivamente. Los datos de entrenamiento se utilizaron para alimentar al modelo y que este pueda analizar y aprender de las correlaciones necesarias dentro de la información para encontrar el mejor camino hacia la estimación de la variable deseada.

3.5 Manejo de Métricas de Evaluación

3.5.1 Aplicación del coeficiente de determinación

Se utilizó el coeficiente de correlación r de Pearson para medir el grado de asociación lineal entre las variables. En este caso se utilizó para calcular la correlación entre las variables seleccionadas “X” para el entrenamiento de la IA y las variables

objetivo “Y” (Nitrógeno y/o Fósforo) y de este modo lograr determinar la capacidad que tiene X para predecir Y. Conceptualmente, r^2 es la fracción de la variabilidad de Y que queda explicada por su dependencia de la variable X. Así, un coeficiente de determinación de 0,64 significa que de la variabilidad total de Y, un 64% se explica por su relación con X y el resto por otros factores desconocidos (Dagnino S., 2014).

Para esto se utilizó la librería de “Scikit-learn” en Python con su modulo específico “sklearn.metrics” y la función “r2_score”. Este indicador proporciona una medida cuantitativa de la calidad de ajuste del modelo, expresando la proporción de variabilidad en la variable dependiente que puede ser explicada por el modelo.

3.5.2 Aplicación de la Raíz del Error Cuadrático Medio (RMSE)

Para calcularlo se utilizó la librería de “Scikit-learn” en Python con su modulo específico “sklearn.metrics” y la función “mean_squared_error”. Se le solicito al algoritmo calcular el MSE para poder obtener el RMSE usando “np.sqrt”, una función de la biblioteca Numpy para calcular la raíz exacta de un número, este caso e utiliza para encontrar la raíz cuadrada del Error Cuadrático Medio (MSE) y así obtener el RMSE. Esto se hace para expresar la métrica en la misma unidad que la variable original y hacerla más interpretable.

3.5.3 Aplicación de la Desviación Absoluta Media

Se calculó utilizando la librería de “Scikit-learn” en Python con su modulo específico “sklearn.metrics” nuevamente pero esta vez haciendo uso de la función “mean_absolute_error”. Para obtener directamente el valor deseado.

4 RESULTADOS

4.1 Evaluación del modelo

Como se mencionó anteriormente, el algoritmo seleccionado fue “Random Forest”. Este algoritmo de aprendizaje automático, desarrollado por Leo Breiman y Adele Cutler, es ampliamente utilizado y se caracteriza por combinar las salidas de varios árboles de decisión para obtener un resultado único. Su popularidad se debe a su facilidad de implementación y flexibilidad, lo que lo hace adecuado tanto para problemas de clasificación como de regresión.

Para dividir la información se utilizó el algoritmo “RandomForest” en 4 modelos que se presentan en la tabla 4-1.

TABLA 4-1 SEPARACIÓN DE CONJUNTO DE DATOS POR TEMPORADA Y SELECCIÓN DE CARACTERÍSTICAS.

Modelo	Temporada de muestreo	Variable objetivo	Variables predictoras
1	Estiaje	Nitrógeno	HG_TOT, ORTO_PO4, P_TOT, COLOR_VER, COLI_TOT, OD, DBO_SOL, TURBIEDAD, TEMP_AGUA, pH, ABS_UV, TEMP_AMB, CR_TOT, COT_SOL, COT, E_COLI, CAUDAL
2	Lluvia	Nitrógeno	ORTO_PO4, TURBIEDAD, SST, DBO_TOT, ABS_UV, P_TOT, DQO_TOT, E_COLI, COLOR_VER, COT_SOL, COT, COLI_FEC, COLI_TOT, AS_TOT, pH, OD, DUR_TOT, TEMP_AMB
3	Estiaje	Fósforo	TURBIEDAD, N_TOTK, PB_TOT, N_NO3, N_TOT, COLOR_VER, ORTO_PO4, N_ORG, DUR_TOT, N_NH3, ABS_UV, COT_SOL, E_COLI, SST, DQO_TOT, COT, OD, CONDUCT, CAUDAL, TEMP_AMB
4	Lluvia	Fósforo	SST, E_COLI, TURBIEDAD, N_TOTK, N_ORG, DQO_TOT, COLOR_VER, DBO_TOT, N_TOT, COLI_FEC, N_NO3, ABS_UV,

COT_SOL, DUR_TOT, N_NH3, CAUDAL, OD_mg/L, OD_%,
pH_CAMPO, TEMP_AMB

La correlación entre dichos datos se presenta en la figura 4-1.

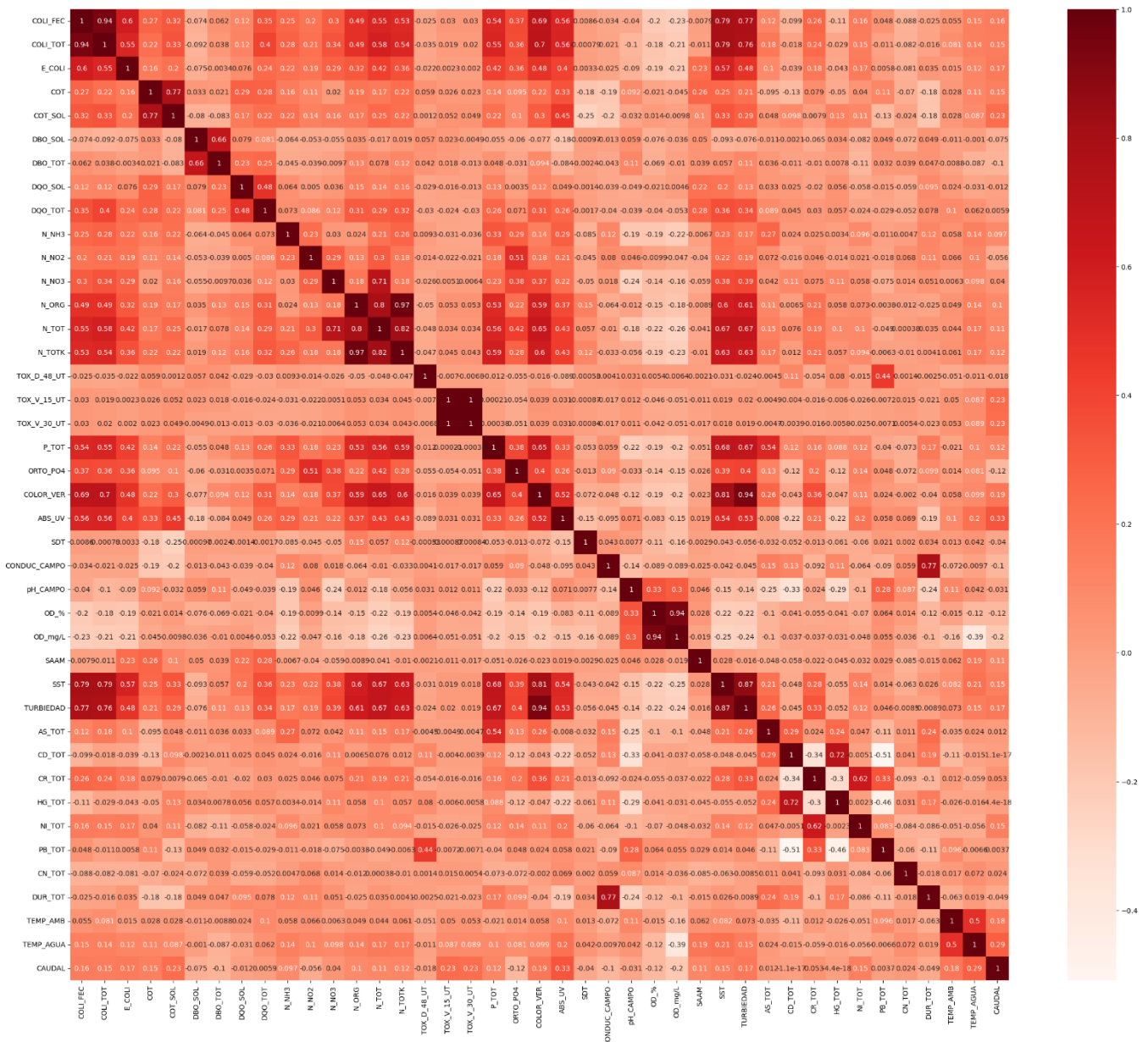


Figura 4-1. Correlación entre parámetros usando la correlación Pearson en Python.

El comportamiento de la gráfica es razonable debido al coeficiente de determinación obtenido con el algoritmo, al imaginar una línea recta intentando atravesar todos los puntos se puede apreciar como en efecto hay una desviación entre los valores

verdaderos y los estimados pero el comportamiento general de los datos si tienen coherencia al seguir de manera lineal.

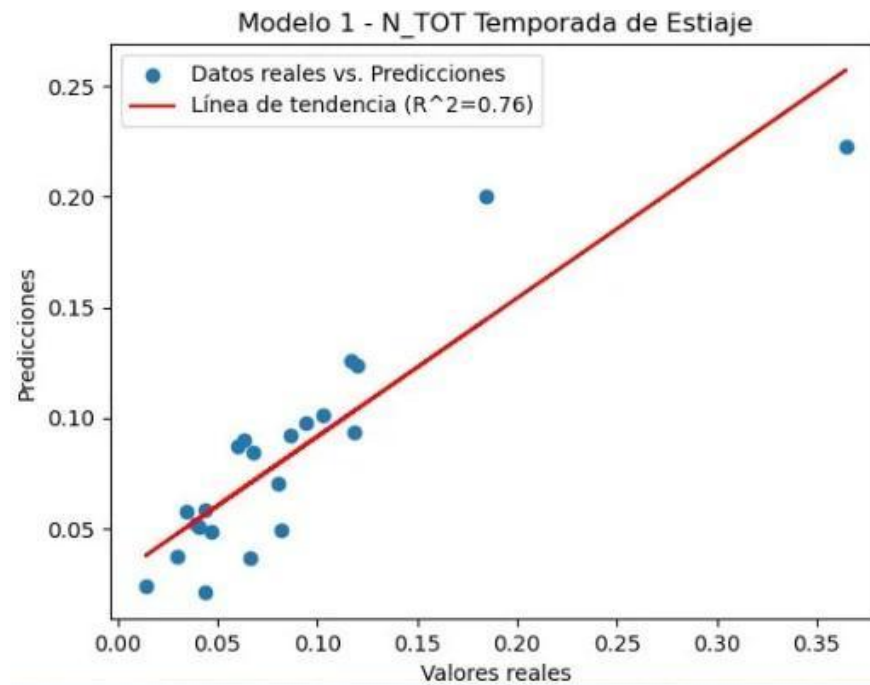


Figura 4-2: Visualización objetiva de los valores predictivos de Nitrogeno Total vs los valores reales en temporada de estiaje.

Como se puede observar todos los modelos cuentan con un buen desempeño en la predicción de la variable objetivo, con un alto coeficiente de determinación y errores relativamente bajos. Aunque es notable que el mejor comportamiento lo tiene el segundo modelo el cual predice Nitrógeno durante la temporada de lluvia. Esto puede apreciarse fácilmente en la Figura 4-3.

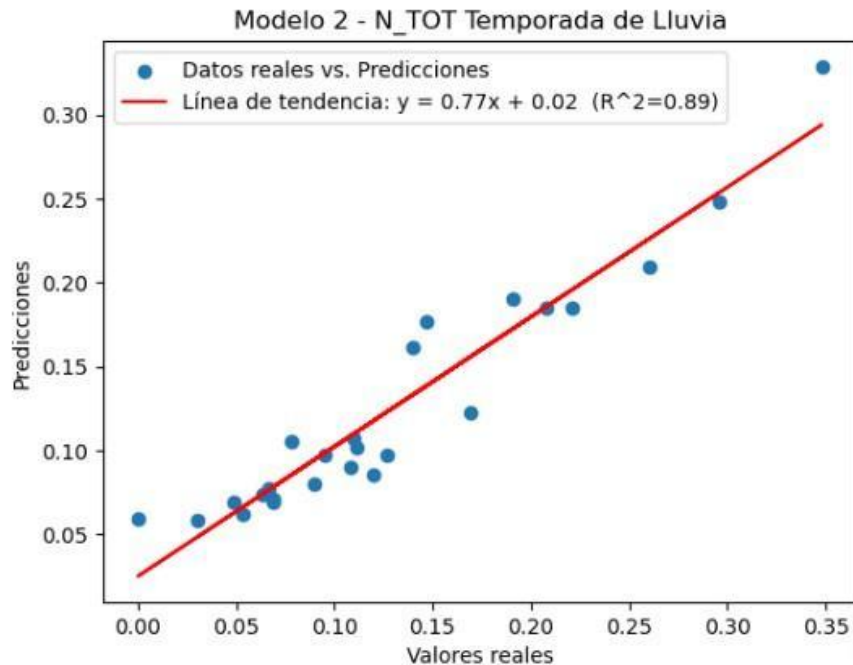


Figura 4-3: Visualización objetiva de los valores predictivos de Nitrogeno Total vs los valores reales en temporada de lluvia.

En la figura 4-3 es fácilmente apreciable como el comportamiento de los datos tiene un desempeño mejor a comparación del resto de modelos, esto puede deberse a una mejor alineación entre la variable objetivo (nitrógeno) y las variables predictoras seleccionadas por el algoritmo.

Mientras que en la figura 4-4 se denota un comportamiento menos lineal, con algunos puntos sobrepuestos y dispersos.

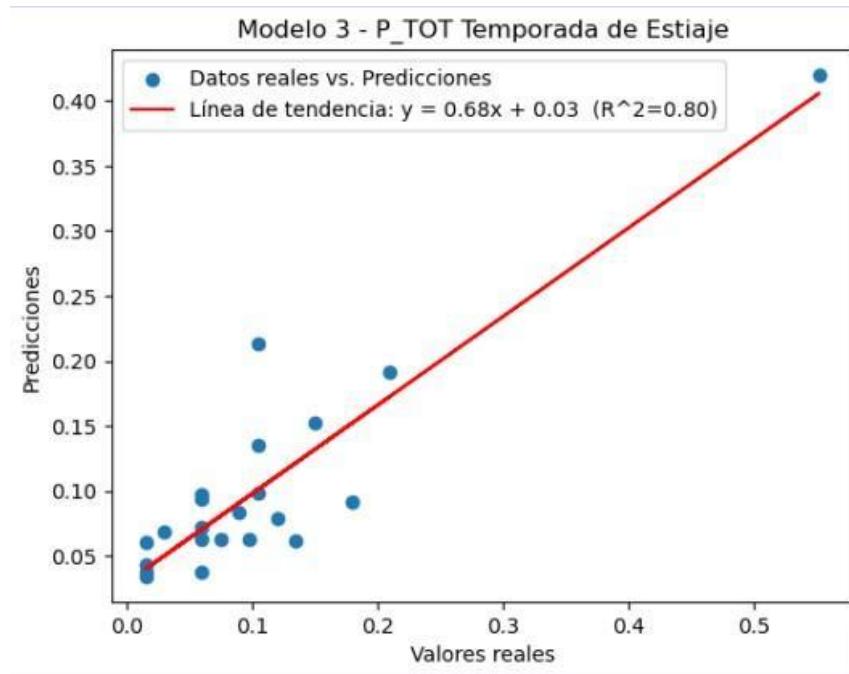


Figura 4-4: Resultado predictivo del Fosforo Total vs los valores reales en temporada de estiaje.

En la figura 4-5 nuevamente se nota un mejor comportamiento en comparación al anterior, por lo que se puede concluir que los datos obtenidos del conjunto de datos formado por los muestreos realizados durante la temporada de lluvia brindan información más completa al modelo a pesar de ser menor cantidad.

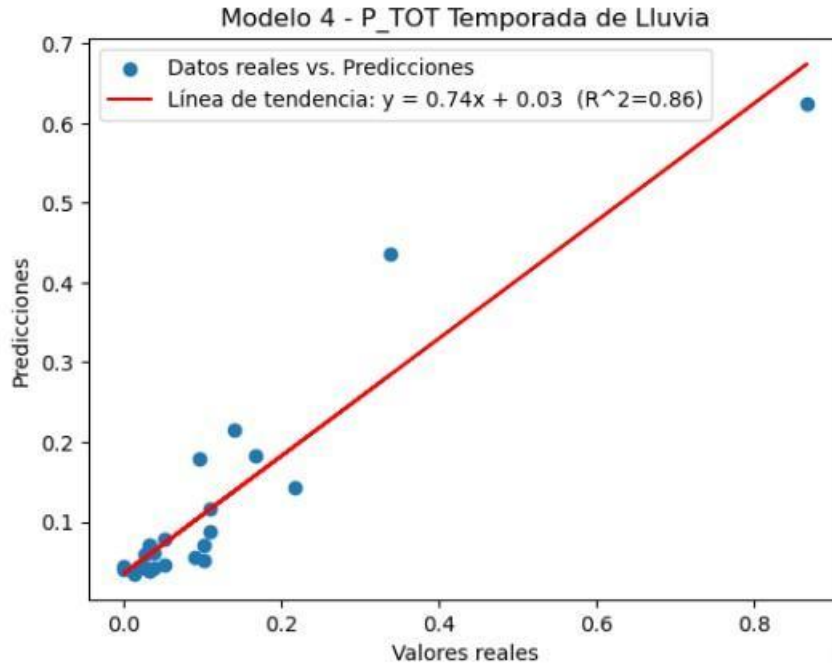


Figura 4-5: Resultado predictivo del Fosforo Total vs los valores reales en temporada de lluvia.

Los 4 conjuntos fueron utilizados para entrenar 4 modelos utilizando RandomForest y para evaluar su confiabilidad se usaron las métricas: Coeficiente de determinación (R^2), Raíz del Error Cuadrático Medio (RMSE) y la Desviación Absoluta Media (MAE), los resultados se muestran en la tabla 4-2.

TABLA 4-2: EVALUACIÓN DE MODELOS.

Modelo	R^2	RMSE	MAE
1	0.76	0.035	0.021
2	0.89	0.060	0.034
3	0.80	0.047	0.033
4	0.860	0.064	0.040

Para comprender mejor estos resultados en la tabla 4-3 se comparan las predicciones con los datos reales de manera directa, mostrando la diferencia numérica entre ambos valores para el caso del modelo 2 que predice nitrógeno durante la temporada de lluvia.

TABLA 4-3: DIFERENCIA NUMÉRICA ENTRE VALORES REALES Y PREDICCIONES EN EL MODELO 2.

Predicciones	Reales	Diferencia
0.087299127	0.108474307	-0.021
0.070185534	0.069159296	0.001
0.0727439	0.066268487	0.006
0.15927845	0.147211156	0.012
0.071312949	0.063377677	0.008
0.100336678	0.111076036	-0.011
0.122783815	0.168892229	-0.046
0.223340628	0.296087853	-0.073
0.083728977	0.119748465	-0.036
0.200156335	0.259952732	-0.060
0.154055335	0.139984132	0.014
0.191429797	0.207918159	-0.016
0.071905565	0.048923629	0.023

Por otro lado, para poder estimar si el cuerpo de agua es considerado como eutrofizado existen distintos índices de estados tróficos aplicados a otros cuerpos de agua con los

cuales se buscaba comparar las cantidades de nutrientes y verificar si el Río Presidio tiene o no un nivel preocupante de esta problemática. Sin embargo, la cantidad de clorofila no fueron medidas en el río presidio (datos faltantes) y es un parámetro importante para realizar cálculos como el Índice del Estado Trófico de Carlson o Índice de Estado Trófico (TRIX).

Por lo que, se realizó una comparativa con las cantidades de fósforo y la normativa vigente actual, donde, de acuerdo con lo que se establece en Criterios Ecológicos de Calidad del Aguas CE-CCA-001/89 en el Diario Oficial de la Federación para la protección de la vida acuática en agua dulce la cantidad máxima permisible de Fosfatos (como PO_4) es de 5 mg/L como se muestra en la figura 4-6.

**Criterios de calidad del agua: niveles máximos en miligramos por litro*
(excepto que se indique otra unidad)**

Parámetro	Fuente de abastecimiento de agua potable	Recreativo con contacto primario	Riego agrícola	Pecuario	Protección de la vida acuática	
					Agua dulce	Agua marina (áreas costeras)
Coliformes fecales (NMP/100 ml)	1000	1	1000	6	1	1
DBO	6	6	6	6	6	6
DQO	6	6	6	6	6	6
Nitratos (como N)	5	6	6	90	6	0.04
Nitrógeno amoniacal	6	6	6	6	0.06	0.01
Oxígeno disuelto ²	4	6	6	6	5	5
Sólidos disueltos totales	500	6	500 ³	1000	6	6
Sólidos suspendidos totales	500	6	50	6	4	4
Fosfatos (como PO_4)	0.1	6	6	6	5	0.002

Figura 4-6: Criterios ecológicos de calidad del agua CE-CCA-001/89 (SEMARNAT, 2018)

Y debido a que en la base de datos proporcionada por CONAGUA no se supera en ninguno de los muestreos desde 2012 el límite señalado en la figura 4-5 se puede concluir que, para la normativa actual, las cantidades de fósforo en forma de fosfatos es normal en esta clase de fluentes. Sin embargo, se considera importante una actualización de la normativa haciendo énfasis en la problemática de la eutrofización para tener un mejor control en la salud de los cuerpos de agua de la república.

5 CONCLUSIONES

Como se menciona en los resultados se concluye que el modelo con mayor capacidad predictora es el segundo, el cual predice las cantidades de nitrógeno en el Río Presidio durante la temporada de lluvia, lo cual es notorio debido a su mayor coeficiente de determinación (0.89) y bajo nivel de error (0.06), esto puede deberse a diversos motivos, desde un mejor ajuste durante el entrenamiento del algoritmo o incluso a una mejor limpieza de datos durante el preprocesamiento, el hecho que este modelo tenga un mejor desempeño que los demás no quiere decir que el restante de modelos no cumplió con el objetivo de predecir las variables asignadas, de hecho, todos los modelos presentaron un buen coeficiente de determinación y demostraron ser aptos para la predicción de variables, ya sea nitrógeno o fosforo en temporadas de lluvia o de estiaje.

Al realizar una comparación entre el modelo con mejor comportamiento (2) contra el que tuvo menor rendimiento de los 4 (1) sobresale el hecho que ambos modelos predicen el mismo nutriente, su diferencia es la base de datos utilizada, donde uno usa los datos durante la temporada de estiaje y el otro usa los datos durante la temporada de lluvia, factor que puede afectar debido a que las concentraciones varían entre estas épocas y dicha diferencia puede generar ruido durante el entrenamiento del algoritmo.

Estos índices demuestran por si mismos que la inteligencia artificial tiene mucho potencial en su utilización para el monitoreo de calidad del agua y puede llegar a ser una gran herramienta para investigadores, dependencias encargadas de monitorear

dicha calidad o incluso empresas que cuenten con plantas de tratamiento. Las posibilidades del uso de esta herramienta en este sector resultan muy diversas y sumamente útiles. Para próximas investigaciones se sugiere llevar esto al siguiente paso, realizar las estimaciones tomando en cuenta la variable tiempo para posibilitar la predicción de datos a un futuro próximo para prevenir posibles brotes de contaminación. U utilizar una cantidad menor de variables predictoras para hacer mas eficiente el proceso de estimación.

En el caso del segundo objetivo de este trabajo, se concluyó que las cantidades de fósforo encontradas en el rio presidio a lo largo de los años no es suficiente para considerarlo una problemática aún y por consecuencia se puede concluir que no se encuentran señales de eutrofización en el cuerpo de agua, sin embargo la normativa no especifica datos acerca del nitrógeno u alguno de sus derivados como nitritos o nitratos, solamente de clorofila, datos que no se encontraban en la base de datos de CONAGUA por lo que se recomienda empezar a realizar este estudio en el río para futuras investigaciones referentes a la eutrofización y posibilitar el uso de los índices disponibles, además de armar un historial de datos de esta variable y poder estudiar su comportamiento con el paso del tiempo.

6 COMPETENCIAS DESARROLLADAS

Las competencias que se desarrollaron a lo largo de la elaboración de esta tesis con título “Estimación de parámetros de calidad del agua en cuenca hidrológica Río presidio con el apoyo de algoritmos de inteligencia artificial” fueron principalmente los conocimientos en un área completamente nueva como lo es la programación de una inteligencia artificial y aplicar dichos conocimientos en un área conocida como lo es la calidad del agua para poder lograr estimaciones de componentes contaminantes en un cuerpo de agua de gran importancia para el estado de Sinaloa, así como la resolución de los problemas presentados durante la investigación y la utilización de los algoritmos y de los análisis ambientales realizados con los datos presentados por CONAGUA.

7 FUENTES DE INFORMACIÓN

1. Díaz-Delgado, C., & Moeller-Chavez, G. (30 de 04 de 2019). *Ingeniería del agua*. Obtenido de Seguridad hídrica en México: diagnóstico general y desafíos principales: <https://iwaponline.com/IA/article/23/2/107/68023/Seguridad-hidrica-en-Mexico-diagnostico-general-y>
2. Santos, D., Dallos, L., & Gaona-García, P. (06 de 2020). *Algoritmos de rastreo de movimiento utilizando técnicas de inteligencia artificial y machine learning*. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-07642020000300023&script=sci_arttext&tlng=pt
3. Aguas Urbanas. (17 de 11 de 2018). *Conceptos sobre monitoreo de calidad de agua*. Obtenido de [http://www.aguasurbanas.ei.udelar.edu.uy/index.php/2018/11/17/conceptos-sobre-monitoreo-de-calidad-de-agua/#:~:text=En%20particular%2C%20un%20monitoreo%20de,acu%C3%A1tica%20\(Chapman%2C%201996\).](http://www.aguasurbanas.ei.udelar.edu.uy/index.php/2018/11/17/conceptos-sobre-monitoreo-de-calidad-de-agua/#:~:text=En%20particular%2C%20un%20monitoreo%20de,acu%C3%A1tica%20(Chapman%2C%201996).)
4. ALARCÓN, M. R. (2023). *ANÁLISIS DE REINCIDENCIA DEL PROGRAMA VOLVER A EMPEZAR UTILIZANDO REGRESIÓN LOGÍSTICA, RANDOM FOREST Y XGBOOST* . Obtenido de UNIVERSIDAD DE CHILE : <https://repositorio.uchile.cl/handle/2250/195252>
5. Aluja, T. (04 de 2001). *La minería de datos, entre la estadística y la inteligencia artificial* . Obtenido de <https://upcommons.upc.edu/bitstream/handle/2099/4162/article.pdf?sequence=4&isAllowed=y>
6. Ambientum . (24 de 11 de 2022). *El nitrógeno y sus compuestos*. Obtenido de Grupo de Tratamiento de Aguas Residuales. Escuela Universitaria Politécnica. Universidad de Sevilla: https://www.ambientum.com/enciclopedia_medioambiental/aguas/consideraciones_del_nitrogeno_sus_compuestos.asp#:~:text=En%20un%20agua%20natural%20se,derivados%20de%20la%20piridina%2C%20etc.

7. Arreguín-Cortés, F. I., López-Pérez, M., & Cervantes-Jaimes, C. E. (27 de 05 de 2020). *Tecnología y ciencias del agua - SciELO*. Obtenido de Los retos del agua en México: https://www.scielo.org.mx/scielo.php?pid=S2007-24222020000200341&script=sci_arttext#B14
8. Aruna Flarence, R. (06 de 2018). *Importance of Supervised Learning in Prediction Analysis*. Obtenido de Periodicals of Engineering and Natural Sciences: : <http://pen.ius.edu.ba>
9. Bem, C. C., Borba Braga, M. C., & Rodrigues de Azevedo, J. C. (01 de 2013). *Avaliação do estado trófico de um lago urbano raso. Revista de Gestão de Água da América Latina*. Obtenido de Asociación Brasileña de Recursos Hídricos : <https://www.abrhidro.org.br/SGCv3/publicacao.php?PUB=2&ID=127&SUMARIO=1699>
10. Cañete, C. (06 de 2019). *La importancia del control y monitoreo de la calidad del agua del Río Paraguay para el desarrollo y la defensa nacional*. Obtenido de <https://doi.org/10.18004/rcfacen.2019.10.1.17>
11. Cárdenas Calvachi, G., & Sánchez Ortiz, I. (13 de 06 de 2007). *Nitrógeno en aguas residuales: orígenes, efectos y mecanismos de remoción para preservar el ambiente y la salud pública*. Obtenido de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0124-71072013000100007
12. Casanova, Y. O. (10 de 2022). *Contribuciones al problema de clasificación en machine learning*. Obtenido de <https://dialnet.unirioja.es/servlet/tesis?codigo=319445>
13. Centro Nacional de Prevención de Desastres . (5 de 06 de 2018). *En esta temporada de lluvias monitorea los niveles de presas, ríos, cauces y arroyos*. Obtenido de <https://www.gob.mx/cenapred/articulos/en-temporada-de-lluvias-monitorea-el-nivel-de-presas-rios-causes-y-arroyos>
14. Chapman, D. V. (1996). *Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring*. .

15. CONAGUA. (2014). *Estadísticas del agua en México*. CDMX: Secretaría de Medio Ambiente y Recursos Naturales.
16. CONAGUA. (2021). *Comisión Nacional del Agua*. Obtenido de Estadísticas del Agua: <https://files.conagua.gob.mx/conagua/publicaciones/Publicaciones/EAM%202021.pdf>
17. CONAGUA. (25 de 10 de 2022). *Estructura de la Red Nacional de Medición de Calidad del Agua (RENAMECA)*. Obtenido de Comisión nacional del agua: <https://app.conagua.gob.mx/ica/Contenido?n1=1&n2=1>
18. CONAGUA. (25 de 10 de 2022). *Estructura de la Red Nacional de Medición de Calidad del Agua (RENAMECA)*. Obtenido de Comisión nacional del agua: <https://app.conagua.gob.mx/ica/Contenido?n1=1&n2=1>
19. CONSELVA. (01 de 06 de 2023). *Conoce tu cuenca: Río Presidio*. Obtenido de http://web.archive.org/web/202300000000000*/https://conselva.org/conoce-tu-cuenca-rio-presidio/
20. Dagnino S., J. (2014). *Coefficiente de correlación lineal de Pearson*. Obtenido de Rev Chil Anest 2014; 43: 150-153: https://www.sachile.cl/upfiles/revistas/54e63a1a778ff_15_correlacion-2-2014_edit.pdf
21. Daily, G., Alexander , S., Ehrlich, P., Goulder, L., Lubchenco, J., Matson, P., . . . Woodwell, G. (1997). *Ecosystem services: benefits supplied to human societies by natural ecosystems*. Issues in Ecology 2.
22. Dominguez, E., Encalada, A., Fernandez, H. R., Giorgi,, A., Marchese Garelo, M. R., Miserendino, M. L., . . . Rodrigues Capitulo, A. (08 de 2021). *Biomonitoreo en ríos de Argentina: un camino por recorrer*. Obtenido de Asociación Argentina de Ecología: <https://ri.conicet.gov.ar/handle/11336/149393>
23. Doyle, M., Stanley, E., Strayer, D., Jacobson, R, & Schmidt, J. (2005). *Effective discharge analysis of ecological processes in streams*. Water Resources Research.

24. Faustino, L. I., Graciano, C., Gortari, F., & Guiamet, J. J. (12 de 2011). *¿Cómo afectan los nutrientes el uso del agua en plantas leñosas?* Obtenido de Asociación Argentina de Ecología.
25. García, C. M., Vicario, L., Colladon, L., Ravelo, A., & Rodríguez, A. (s.f.). *INCERTIDUMBRE EN LA ESTIMACIÓN DE CAUDALES DE PERÍODOS DE ESTIAJE*. Obtenido de Instituto Superior de Recursos Hídricos: http://irh-fce.unse.edu.ar/Rios2007/index_archivos/D/10.pdf
26. Google Earth. (07 de 12 de 2022). *Google*. Obtenido de <https://earth.google.com/web/search/R%c3%ado+Presidio/@23.60944146,-106.0696874,261.58307881a,12291.51408346d,35y,0h,0t,0r/data=CngaThJlCiUweDg2OThlMWI3MGI2MjkyZGY6MHg4ZTZkYTI5MTg1NzY0GRVFWQknnTdAlcNqhxrpg1rAKg1Sw61vIFByZXNpZGlvGAlGASlmcjQJ6PF5AmBcOEA>
27. Hispagua Sistema Español de Información sobre el Agua. (06 de 2007). *Energía Hidraulica*. Obtenido de https://hispagua.cedex.es/sites/default/files/especiales/energia_hidr/4_impacto.htm#:~:text=La%20construcci%C3%B3n%20de%20presas%20y,y%20mayor%20proliferaci%C3%B3n%20de%20enfermedades.
28. INEGI. (25 de 04 de 2019). *Clima*. Obtenido de <https://cuentame.inegi.org.mx/monografias/informacion/dur/territorio/clima.aspx#:~:text=Durango&text=El%2040%25%20del%20territorio%20presenta,es%20de%2017%20C2%B0C>.
29. INEGI. (23 de 10 de 2023). *Clima*. Obtenido de <https://cuentame.inegi.org.mx/monografias/informacion/sin/territorio/clima.aspx?tema=me&e=25>
30. INEGI. (06 de 2024). *Usos del agua en México*. Obtenido de <https://cuentame.inegi.org.mx/territorio/agua/usos.aspx?tema=T#:~:text=En%20M%C3%A9xico%20C%2076%20%25%20del%20agua,5%20%25%20C%20en%20la%20industria>.

31. Larrea Murrell, J., Romeu Alvarez, B., Lugo Moya, D., & Rojas Badía, M. (05 de 2022). *Aspectos fundamentales del monitoreo de calidad de las aguas: el río almendares como caso de estudio*. Obtenido de http://scielo.sld.cu/scielo.php?pid=S2221-24502022000200148&script=sci_arttext

32. Lozano, M. G. (26 de 09 de 2016). *Eutrofización: una visión general*. Obtenido de <http://www.cienciacierta.uadec.mx/2016/09/26/eutrofizacion-una-vision-general/>

33. Lupi, O. D., Canziani, M. B., & Zaradnik, I. J. (12 de 2020). *ESTADO DEL ARTE DE LOS SISTEMAS DE*. Obtenido de <https://repositoriocyt.unlam.edu.ar/bitstream/123456789/1221/1/ReDDi%205-2-5.%20Estado%20de%20arte%20de%20los%20sistemas%20de%20monitoreo%20de%20calidad%20de%20agua%20.pdf>

34. Martín-Romero, A. C. (04 de 2020). *Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas*. Obtenido de <https://repositorio.comillas.edu/xmlui/handle/11531/45878>

35. Mendiondo, E., Clarke, R., & Tönsmann, F. (2000). *River restoration, discharge uncertainties and floods*. Alemania: F. Tönsmann & M. Koch "River Flood Defence", Herkules Vg, Kassel.

36. Ministerio para la Transición Ecológica y el Reto Demográfico. (s.f.). *Gases eutrofizantes*. Obtenido de https://www.miteco.gob.es/en/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/emisiones/prob-amb/gases_eutrofizantes.html

37. Monforte García, G., & Cantú Martínez, P. (01 de 2009). *ESCENARIO DEL AGUA EN MEXICO*. Obtenido de https://www.researchgate.net/publication/44960104_Escenario_del_agua_en_Mexico

38. Parra, F. (25 de 01 de 2019). *Estadística y Machine Learning con R*. Obtenido de <https://bookdown.org/content/2274/metodos-de-clasificacion.html>

39. Patiño Pérez, D., Silva Bustillos, R., Munive Mora, C., & Botto Tobar, M. (15 de 09 de 2020). *Predicción de Covid19 con el uso del Algoritmo Random Forest y Redes Neuronales Artificiales*. Obtenido de <http://portal.amelica.org/ameli/jatsRepo/606/6062588014/index.html>
40. Postel, S., & Brian, R. (2003). *Rivers for life. Managing water for people and nature*. EE.UU.: Island Press.
41. Riabani Mercado, F., García Fernández, W., & Herrera Acebey, J. (1 de 09 de 2016). *Sistema de inteligencia artificial para la predicción temprana de heladas meteorológicas*. Obtenido de http://www.scielo.org.bo/scielo.php?pid=S1683-07892016000200007&script=sci_arttext
42. Rodríguez-Lara, J., Cervantes-Ortiz, F., Arámbula-Villa, G., Mariscal-Amaro, L., Aguirre-Mancilla, C., & Andrio-Enríquez, E. (2022). *Lirio acuático (Eichhornia crassipes): una revisión*. Obtenido de Universidad de Costa Rica: <https://revistas.ucr.ac.cr/index.php/agromeso/index>
43. Salas, R. (2004). *Redes Neuronales Artificiales*. Obtenido de Universidad de Valparaíso: https://www.academia.edu/download/50358783/Redes_Neuronales_Artificiales.pdf
44. SEMARNAT. (1998). *NOM-003-ECOL-1997*. Obtenido de Que establece los límites máximos permisibles de contaminantes para las aguas residuales tratadas que se reusen en servicios al público.: https://diariooficial.gob.mx/nota_detalle.php?codigo=4863357&fecha=14/01/1998&print=true
45. SEMARNAT. (2022). *NOM-001-SEMARNAT-2021*. Obtenido de Que establece los límites permisibles de contaminantes en las descargas de aguas residuales en cuerpos receptores propiedad de la nación.: https://www.dof.gob.mx/nota_detalle.php?codigo=5645374&fecha=11/03/2022#gsc.tab=0
46. SEMARNAT. (04 de 06 de 2023). Obtenido de INDICADORES DE ESTADO: https://apps1.semarnat.gob.mx:8443/dgeia/indicadores14_cd/conjuntob/02_agua/02_calidad_estado.html#:~:text=El%20f%C3%B3sforo%20en%20el%20agua,la%20descomposici%C3%B3n%20de%20%C3%A9stas%20cuando

47. Sukhdev, P. (2018). *WWF Internacional*. Obtenido de https://www.wwf.org.mx/que_hacemos/agua_dulce/valorar_los_rios/#:~:text=Durante%20siglos%2C%20se%20han%20usado,bienestar%20econ%C3%B3mico%20en%20las%20comunidades.
48. Sultenfuss, J., & Doyle, W. (1999). Functions of Phosphorus in Plants. *Better Crops with plant food*, 6.
49. Timón, C. E. (16 de 01 de 2017). *“Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso*. Obtenido de <https://openaccess.uoc.edu/bitstream/10609/59565/6/caresptimTFG0117mem%C3%B2ria.pdf>
50. Ulrich, K., & Eppinger, S. (2013). *Diseño y desarrollo de productos* (5ª edición ed.). México, D.F.: Mc Graw Hill.
51. UNEP. (23 de 09 de 2022). *Cuatro razones para proteger los ríos*. Obtenido de <https://www.unep.org/es/noticias-y-reportajes/reportajes/cuatro-razones-para-proteger-los-rios#:~:text=Los%20r%C3%ADos%20sustentan%20a%20las%20personas%20y%20la%20econom%C3%ADa&text=Mientras%20tanto%2C%20los%20r%C3%ADos%20proporcionan,cuales%20el%2055%25%25>
52. UNEP. (23 de 09 de 2022). *ONU programa para el medio ambiente* . Obtenido de Cuatro razones para proteger los ríos: <https://www.unep.org/es/noticias-y-reportajes/reportajes/cuatro-razones-para-proteger-los-rios#:~:text=Los%20r%C3%ADos%20sustentan%20la%20biodiversidad&text=Los%20nutrientes%20transportados%20por%20los,agua%20dulce%2C%20seg%C3%BAn%20el%20WWF.>
53. Vide, J. P. (12 de 2006). *Ingeniería de ríos*. Obtenido de https://books.google.es/books?hl=es&lr=&id=DIDodO5iHEYC&oi=fnd&pg=PT6&dq=rios&ots=Fu5u_BA-b7&sig=TWEK6-Zd5UFDjklGbqS_DUJIWWo#v=onepage&q=rios&f=false

8 REGISTRO DE PRODUCTOS

8.1 Artículo publicado en revista de divulgación científica

Se cuenta con la publicación en la revista “Ciencia y Naturaleza” de un artículo con título “Inteligencia artificial como alternativa para la calidad del agua” y fecha de publicación el 9 de septiembre del 2024.

Se encuentra disponible en: <https://www.revistacyn.com/actual/1108>



Figura 8-1 Carta de aceptación de contribución a revista de divulgación “Ciencia y Naturaleza”

8.2 Participación como expositora en congreso internacional

Exposición de trabajo con título “ESTIMACIÓN DE NUTRIENTES A TRAVÉS DE INTELIGENCIA ARTIFICIAL PARA EL MONITOREO DE CALIDAD DEL AGUA EN CUENCAS AGRÍCOLAS” En el área de INGENIERÍA AMBIENTAL en modalidad ORAL en el XLV encuentro nacional de la Academia Mexicana de Investigación y Docencia en Ingeniería Química, La Inteligencia

Artificial en la Ingeniería Química: Pasado, Presente y Futuro, realizado en Ixtapa–Zihuatanejo, Guerrero, México del 07 al 10 de mayo de 2024.



Figura 8-2. Reconocimiento por presentación de trabajo en congreso “AMIDIQ”

8.3 Presentación de cartel en el Primer Congreso de Investigadoras Sinaloenses

Exposición de cartel en el Primer Congreso de Investigadoras Sinaloenses con el tema “Monitoreo de calidad del agua con inteligencia artificial”, el cual se llevó a cabo el 14 y 15 de marzo del 2024 con el objetivo de fomentar la participación y el empoderamiento de las mujeres en el ámbito de la investigación.



Culiacán Rosales, Sinaloa; 6 de marzo de 2024
ASUNTO: ACEPTACIÓN DE CARTEL.

Lic. Damaris Alejandra Zamudio Trujillo
Presente.-

Me es grato informarle que su proyecto para presentación de cartel "Monitoreo de calidad del agua con inteligencia artificial", ha sido aceptado para presentarse en el Primer Congreso de Investigadoras Sinaloenses, el cual es una iniciativa de la Coordinación General para el Fomento a la Investigación Científica e Innovación del Gobierno del Estado de Sinaloa con el propósito de generar un espacio para la expresión y difusión del trabajo que realizan las investigadoras en Sinaloa y sus aportaciones para el empoderamiento de las mujeres en todas las áreas del conocimiento y en coordinación con el Centro de Políticas de Género para la Igualdad entre Mujeres y Hombres de la Universidad Autónoma de Sinaloa.

Esperando contar con su presencia, le envío un cordial saludo.

Atentamente


Dr. Carlos Karam Quiñones
Director General de CONFIE.

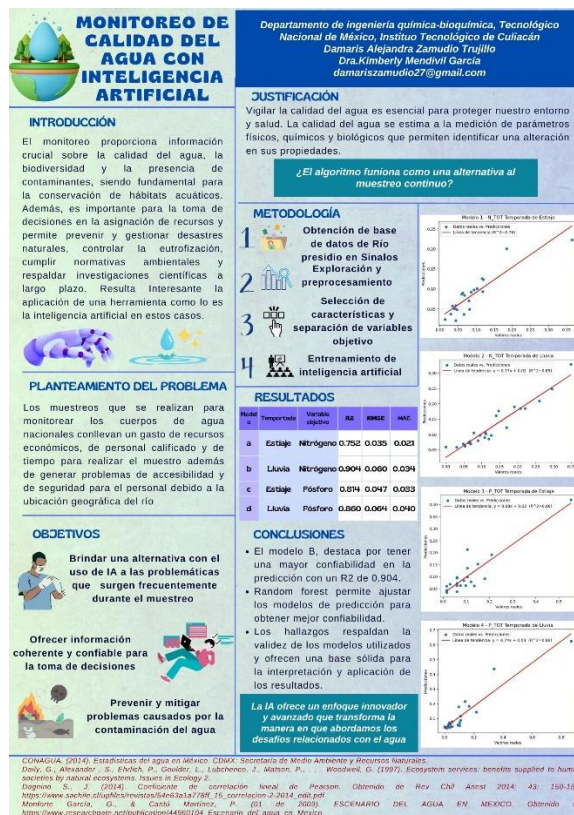


Figura 8-3. Carta de aceptación a presentación de cartel en el Primer Congreso de Investigadoras Sinaloenses y cartel presentado