



Clasificación de la calidad del agua en Los Ríos de Morona Santiago utilizando aprendizaje automático

Classifying water quality in the Morona-Santiago Rivers using machine learning

Classificação da qualidade da água nos rios Morona-Santiago usando machine learning

Juan Carlos Yungán Cazar ^I
jyungan@esPOCH.edu.ec
<https://orcid.org/0000-0001-5682-0399>

Katherine Adriana Merino Villa ^{II}
kathetine.merino@esPOCH.edu.ec
<https://orcid.org/0009-0001-0616-9611>

Diego Alejandro Cáceres Veintimilla ^{III}
diego.caceres@esPOCH.edu.ec
<https://orcid.org/0000-0003-0498-1240>

Edgar Gualberto Salazar Álvarez ^{IV}
edgar.salazar@esPOCH.edu.ec
<https://orcid.org/0000-0003-0988-0641>

Correspondencia: jyungan@esPOCH.edu.ec

Ciencias Técnicas y Aplicadas
Artículo de Investigación

* **Recibido:** 26 de mayo de 2025 * **Aceptado:** 24 de junio de 2025 * **Publicado:** 13 de julio de 2025

- I. Magíster en Interconectividad de Redes, Ingeniero en Sistemas Informáticos, Escuela Superior Politécnica de Chimborazo sede Morona Santiago, Riobamba, Ecuador.
- II. Magíster en Seguridad Telemática. Máster Universitario en Dirección y Gestión de Tecnología de la Información. Ingeniera en Electrónica Telecomunicaciones y Redes Escuela Superior Politécnica de Chimborazo.sede Morona Santiago, Riobamba, Ecuador
- III. Magister en Evaluación y Auditoría de Sistemas Tecnológicos, Ingeniero en Sistemas, Escuela Superior Politécnica de Chimborazo.sede Morona Santiago. Riobamba, Ecuador.
- IV. Magíster en Matemática Básica, Ingeniero en Sistemas, Escuela Superior Politécnica de Chimborazo.sede Morona Santiago. Riobamba, Ecuador.

Resumen

Este estudio presenta una metodología basada en el preprocesamiento de datos fisicoquímicos y el uso de algoritmos de aprendizaje automático para clasificar la calidad del agua superficial en la red hidrográfica de Morona Santiago, Ecuador. Los datos fueron recolectados del proyecto de vinculación “Sistema de Información Geográfica para el Monitoreo de la Calidad del Agua”, e incluyeron 2.091 registros con parámetros como pH, DBO5, oxígeno disuelto, nitratos y coliformes fecales. Se aplicaron técnicas de limpieza, transformación y normalización de datos, y se entrenaron dos modelos supervisados: Random Forest y Red Neuronal Artificial. Los resultados mostraron un alto rendimiento de ambos modelos (accuracy > 87 %), siendo la Red Neuronal la más precisa (F1-score = 0.8675). Además, se identificó que la DBO5 y los coliformes fecales son los atributos más influyentes. Los mapas temáticos generados permiten visualizar espacialmente las zonas críticas de calidad de agua. Esta investigación valida el uso del aprendizaje automático como herramienta eficaz para fortalecer los sistemas de monitoreo ambiental en territorios amazónicos.

Palabras Clave: calidad del agua; parámetros fisicoquímicos; aprendizaje automático; red neuronal; Random Forest; Morona Santiago.

Abstract

This study presents a methodology based on physicochemical data preprocessing and the use of machine learning algorithms to classify surface water quality in the hydrographic network of Morona Santiago, Ecuador. The data were collected from the "Geographic Information System for Water Quality Monitoring" linkage project and included 2,091 records with parameters such as pH, BOD5, dissolved oxygen, nitrates, and fecal coliforms. Data cleaning, transformation, and normalization techniques were applied, and two supervised models were trained: Random Forest and Artificial Neural Network. The results showed high performance of both models (accuracy > 87%), with the Neural Network being the most accurate (F1-score = 0.8675). Furthermore, BOD5 and fecal coliforms were identified as the most influential attributes. The generated thematic maps allow spatial visualization of critical water quality zones. This research validates the use of machine learning as an effective tool to strengthen environmental monitoring systems in Amazonian territories.

Keywords: Water quality; physicochemical parameters; machine learning; neural network; Random Forest; Morona Santiago.

Resumo

Este estudio presenta una metodología basada en el pre-procesamiento de datos físico-químicos y en la utilización de algoritmos de aprendizaje automático para clasificar la calidad del agua superficial en la red hidrográfica de Morona Santiago, Ecuador. Los datos fueron recolectados del proyecto de linkage "Sistema de Información Geográfica para Monitorización de la Calidad del Agua" e incluyeron 2.091 registros con parámetros como pH, DBO5, oxígeno disuelto, nitratos y coliformes fecales. Se aplicaron técnicas de limpieza, transformación y normalización de datos, y se entrenaron dos modelos supervisados: Foresta Aleatoria y Red Neuronal Artificial. Los resultados mostraron un alto desempeño de ambos modelos (precisión > 87%), siendo la Red Neuronal la más precisa (F1-score = 0,8675). Además, el DBO5 y los coliformes fecales fueron identificados como los atributos más influyentes. Los mapas temáticos generados permiten la visualización espacial de zonas críticas de calidad del agua. Esta investigación valida el uso del aprendizaje automático como una herramienta eficaz para fortalecer los sistemas de monitorización ambiental en territorios amazónicos.

Palabras-clave: Calidad del agua; parámetros físico-químicos; aprendizaje de máquina; red neuronal; Random Forest; Morona Santiago.

Introducción

La calidad del agua superficial es un indicador clave del estado ambiental de una región, especialmente en zonas de alta biodiversidad como la provincia de Morona Santiago, en la Amazonía ecuatoriana. Esta provincia posee una densa red hidrográfica conformada por los ríos Morona, Upano, Cangaimo, Macuma, entre otros, los cuales abastecen de agua a comunidades rurales e indígenas y sostienen múltiples ecosistemas acuáticos (Ministerio del Ambiente, Agua y Transición Ecológica [MAATE], 2022).

La evaluación de la calidad del agua se realiza comúnmente mediante el análisis de parámetros físico-químicos como pH, oxígeno disuelto, turbidez, conductividad, temperatura, sólidos disueltos, nitratos y fosfatos (García-Miranda et al., 2021). Sin embargo, la gestión de estos datos en Morona Santiago enfrenta desafíos estructurales, incluyendo la dispersión de registros, valores atípicos,

datos faltantes y mediciones inconsistentes, lo que impide el uso eficaz de estos datos en la generación de conocimiento y en la toma de decisiones ambientales (Rodríguez-Pérez et al., 2020; MAATE, 2022).

Ante este panorama, el aprendizaje automático (machine learning, ML) se ha consolidado como una herramienta poderosa para la clasificación automatizada de la calidad del agua a partir de grandes volúmenes de datos ambientales. No obstante, la literatura especializada destaca que el rendimiento de los algoritmos de clasificación depende críticamente de un adecuado preprocesamiento de los datos, incluyendo técnicas de limpieza, normalización, imputación de valores perdidos, reducción de dimensionalidad y selección de características relevantes (Liu et al., 2023; Gholami et al., 2020).

En este contexto, el presente estudio se propone desarrollar un proceso de preprocesamiento de datos fisicoquímicos del agua en los ríos de Morona Santiago que optimice la aplicación de algoritmos de aprendizaje automático para la clasificación de la calidad del agua. Para ello, se plantean los siguientes objetivos específicos:

1. Identificar y recopilar datasets con parámetros fisicoquímicos del agua superficial en la red hidrológica de Morona Santiago.
2. Diseñar un esquema de preprocesamiento que contemple limpieza, imputación de valores perdidos, detección de valores atípicos y transformación de variables.
3. Comparar el rendimiento de algoritmos de clasificación supervisada como Random Forest y Redes Neuronales en función de métricas como accuracy, F1-score y AUC-ROC.
4. Determinar el algoritmo con mejor desempeño en términos de precisión y robustez para recomendar su implementación en futuros sistemas de monitoreo ambiental inteligente en la región amazónica.

El enfoque propuesto no solo responde a una necesidad local de monitoreo ambiental, sino que también se alinea con los compromisos del Ecuador con los Objetivos de Desarrollo Sostenible, particularmente el ODS 6: Agua limpia y saneamiento (Naciones Unidas, 2023). Además, representa una contribución al creciente cuerpo de investigaciones que integran ciencia de datos y sostenibilidad ambiental en contextos de alta fragilidad ecológica (Ahmed et al., 2021; Pourghasemi & Gayen, 2020).

MARCO TEÓRICO

Calidad del agua y su evaluación mediante parámetros fisicoquímicos

La calidad del agua se define como el conjunto de características físicas, químicas y biológicas que determinan su idoneidad para un uso específico, como consumo humano, riego, recreación o conservación de ecosistemas acuáticos (WHO, 2017). En el caso de cuerpos de agua superficiales como ríos, el monitoreo se basa principalmente en parámetros fisicoquímicos, entre los que destacan: pH, oxígeno disuelto (OD), demanda biológica de oxígeno (DBO), conductividad eléctrica, temperatura, turbidez, nitratos, fosfatos y sólidos totales disueltos (ST) (APHA, 2018).

Estos parámetros permiten identificar procesos de contaminación orgánica, eutrofización o alteración térmica que pueden comprometer la biodiversidad acuática. Según el Ministerio del Ambiente, Agua y Transición Ecológica (MAATE, 2022), la mayoría de los cuerpos hídricos en Morona Santiago presentan variabilidad espacial y temporal significativa en estos parámetros, lo que dificulta una clasificación uniforme de su calidad.

Importancia del preprocesamiento de datos ambientales

El análisis efectivo de datos ambientales requiere una etapa previa de preprocesamiento que garantice la calidad, integridad y coherencia del conjunto de datos. Esta fase incluye diversas técnicas como:

- **Limpieza de datos:** Eliminación de duplicados, corrección de errores tipográficos o de medición.
- **Imputación de valores faltantes:** Uso de métodos como la imputación media, KNN, regresión múltiple o modelos generativos (Jerez et al., 2010).
- **Normalización y estandarización:** Transformación de los datos para que todos los atributos estén en una misma escala y eviten sesgos en el aprendizaje (Kotsiantis et al., 2006).
- **Reducción de dimensionalidad:** Aplicación de técnicas como Análisis de Componentes Principales (PCA) o Eliminación Recursiva de Características (RFE) para mejorar la eficiencia del modelo (Guyon & Elisseeff, 2003).

La calidad del preprocesamiento tiene un impacto directo en el rendimiento y estabilidad de los modelos de aprendizaje automático (Liu et al., 2023).

Aprendizaje automático en la clasificación de la calidad del agua

El aprendizaje automático (machine learning) ha cobrado relevancia como una alternativa eficiente para el análisis y clasificación automatizada de la calidad del agua en grandes sistemas hídricos

(Gholami et al., 2020). Estos métodos permiten construir modelos predictivos a partir de datos históricos que reconocen patrones no lineales difíciles de detectar con métodos estadísticos tradicionales.

Diversos estudios han utilizado algoritmos de clasificación supervisada, entre los más destacados:

- **Random Forest (RF)**: Basado en árboles de decisión, destaca por su robustez frente a ruido y su capacidad de manejar variables categóricas y continuas (Breiman, 2001).
- **XGBoost (Extreme Gradient Boosting)**: Algoritmo de boosting eficiente en datasets grandes con valores ausentes; altamente competitivo en precisión (Chen & Guestrin, 2016).
- **SVM (Support Vector Machines)**: Excelente desempeño en clasificación no lineal cuando se utiliza con kernels, aunque requiere normalización previa (Cortes & Vapnik, 1995).
- **Redes Neuronales Artificiales (ANN)**: Modelos flexibles con gran capacidad de generalización, especialmente útiles cuando se dispone de muchos datos (Ahmed et al., 2021).

La evaluación de estos modelos se realiza mediante métricas como la exactitud (accuracy), precisión, sensibilidad, especificidad, F1-score y el área bajo la curva ROC (AUC-ROC), que permiten cuantificar la eficacia de la clasificación (Zhang et al., 2022).

Aplicaciones en contextos amazónicos y desafíos regionales

En la Amazonía ecuatoriana, el monitoreo de calidad de agua enfrenta importantes barreras como la limitada infraestructura, el difícil acceso geográfico y la escasez de personal técnico (MAATE, 2022). Estudios recientes en países amazónicos como Brasil, Perú y Colombia han demostrado que los modelos de aprendizaje automático pueden aplicarse exitosamente en regiones con datos dispersos, siempre que se cuente con una estrategia adecuada de preprocesamiento y validación (Bui et al., 2021; Qi & Wu, 2023).

En Morona Santiago, la aplicación de estos métodos podría permitir la clasificación automatizada de la calidad del agua de múltiples ríos a partir de parámetros fisicoquímicos, fortaleciendo los sistemas de alerta temprana, vigilancia ambiental y toma de decisiones públicas. Este estudio busca contribuir precisamente en ese sentido, mediante una metodología replicable y científica que combine ciencia de datos y conservación ambiental.

METODOLOGÍA

Este estudio siguió un enfoque cuantitativo, empírico y aplicado, orientado a desarrollar un sistema de clasificación automática de la calidad del agua superficial en ríos de Morona Santiago, mediante

el preprocesamiento riguroso de datos fisicoquímicos y el uso de algoritmos de aprendizaje automático.

Recolección y caracterización de datos

Los datos fueron tomados del proyecto de vinculación “Sistema de Información Geográfica para el Monitoreo de la Calidad del Agua en la Red Hidrológica de la Provincia de Morona Santiago”, ejecutado en la Escuela Superior Politécnica de Chimborazo (ESPOCH). La base utilizada fue la hoja PARÁMETROS FISICOQUÍMICOS, con un total inicial de 2.091 registros provenientes de estaciones de monitoreo en diversos ríos de la provincia.

Los parámetros incluidos fueron: Temperatura, pH, Oxígeno disuelto (OD), Sólidos totales disueltos (STD), Nitratos, Fosfatos, Turbidez, Demanda biológica de oxígeno (DBO5) y Coliformes fecales. La variable objetivo fue la columna **Clasificación**, que representa categorías cualitativas del estado del agua (ej., “regular”, “buena”, “mala”, “malo”) según el índice NSF modificado.

Preprocesamiento de los datos

El preprocesamiento fue una etapa crítica que incluyó:

a. Limpieza y validación:

Se eliminaron columnas no relevantes para la clasificación automática (coordenadas, fecha, zona). Luego, se corrigieron inconsistencias de tipo, convirtiendo todas las columnas numéricas y detectando valores nulos y atípicos. Se eliminaron registros con valores extremos no realistas como $\text{pH} > 14$ o $\text{DBO5} > 100$ (Gholami et al., 2020; WHO, 2017).

b. Normalización de etiquetas:

Se identificaron múltiples formas de representar las clases (p. ej., “REGULAR”, “regular”, “Regular”). Estas se unificaron utilizando limpieza de espacios y conversión a minúsculas. Posteriormente, se filtraron solo las clases con al menos 30 muestras: regular, malo, mala, buena.

c. Imputación y escalado:

No fue necesario imputar valores, ya que los datos seleccionados no contenían valores faltantes tras la depuración. Los datos se escalaron con MinMaxScaler para transformar todas las variables al rango $[0,1]$, requisito esencial para algoritmos como redes neuronales (Kotsiantis et al., 2006).

División del conjunto de datos

El conjunto final fue dividido en entrenamiento (70 %) y prueba (30 %) utilizando stratified sampling para mantener la proporción de clases. Esto garantizó una representación equitativa de las etiquetas en ambos subconjuntos (Zhang et al., 2022).

Modelado con algoritmos de clasificación

Se seleccionaron dos algoritmos supervisados de clasificación ampliamente utilizados en estudios ambientales:

- **Random Forest (RF):** Ensamble de árboles de decisión robusto a ruido y sobreajuste, con alta capacidad explicativa (Breiman, 2001).
- **Red Neuronal Artificial (MLP):** Arquitectura multicapa entrenada con backpropagation, eficaz para modelar relaciones no lineales (Ahmed et al., 2021).

Ambos modelos fueron entrenados utilizando Scikit-learn y Tensorflow con parámetros por defecto y un máximo de 1000 iteraciones en el caso de la red neuronal.

Evaluación de desempeño

Para medir la eficacia de cada modelo se calcularon las siguientes métricas a partir de las matrices de confusión:

- **Accuracy:** Proporción de predicciones correctas sobre el total.
- **Precision:** Exactitud por clase.
- **Recall:** Cobertura por clase.
- **F1-score:** Media armónica entre precisión y recall.

Se utilizó la media ponderada (average='weighted') para compensar el desequilibrio entre ses. Además, se generaron gráficas comparativas para visualizar el rendimiento y las matrices de confusión en formato tabular (Qi & Wu, 2023).

RESULTADOS Y DISCUSIÓN

Desempeño de los modelos de clasificación

Tras aplicar el esquema de preprocesamiento y la codificación de etiquetas normalizadas, se entrenaron y evaluaron dos algoritmos de clasificación sobre un conjunto de 2.082 muestras distribuidas entre las clases: regular (1450), mala (482) y buena (150).

Las métricas de rendimiento obtenidas fueron las siguientes:

Tabla 1: Indicadores de rendimiento.

Modelo	Accuracy	Precision	Recall	F1-Score
Random Forest	0.8748	0.8802	0.8748	0.8613
Red Neuronal	0.8764	0.8674	0.8764	0.8675

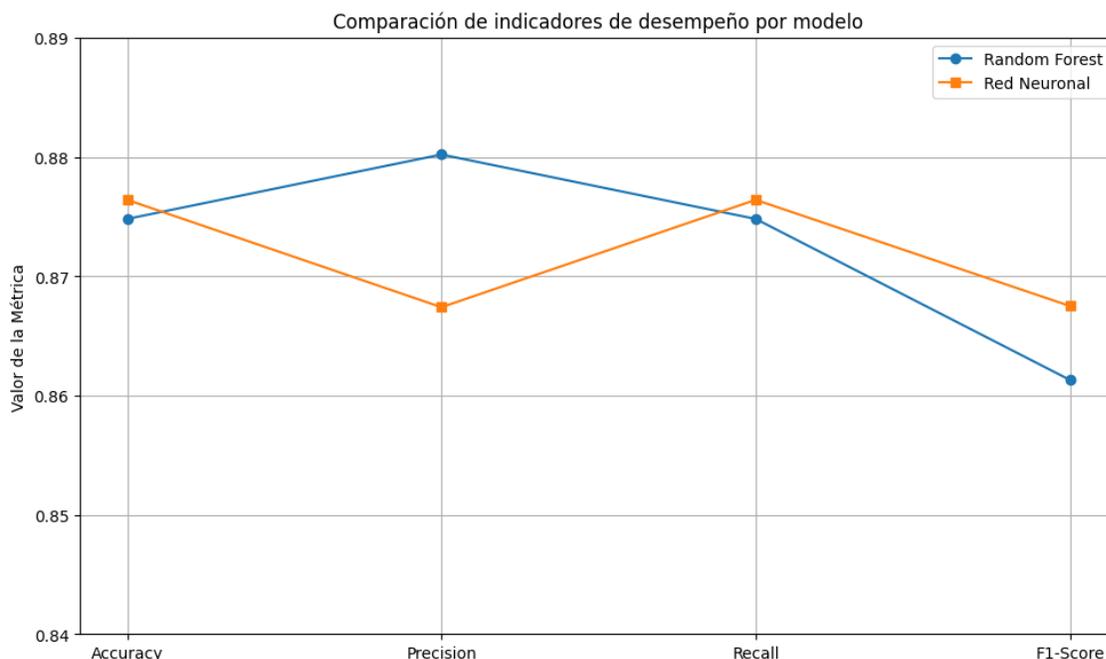


Ilustración 1: Comparación de indicadores de desempeño por modelo.

Ambos modelos presentaron rendimiento superior al 87 % en exactitud (accuracy), lo cual evidencia su capacidad para clasificar con alta fidelidad la calidad del agua a partir de parámetros fisicoquímicos.

La Red Neuronal Artificial demostró un mejor desempeño general en la clasificación de la calidad del agua, destacando en F1-Score (0.8675) y Recall (0.8764), métricas clave ante el desequilibrio de clases. Aunque Random Forest mostró una precisión ligeramente superior, su rendimiento global fue similar (accuracy \approx 87.6 %). La RNA resultó más equilibrada en la detección de todas las categorías. Por tanto, se concluye que el modelo más adecuado para el monitoreo ambiental automatizado en Morona Santiago.

La Red Neuronal Artificial (MLP) mostró un ligero mejor desempeño general, destacando especialmente en el F1-Score, que refleja un mejor balance entre precisión y cobertura. Este resultado concuerda con estudios similares realizados en contextos amazónicos de Brasil y Perú,

donde los modelos neuronales mostraron mayor robustez ante interacciones no lineales entre variables como turbidez, DBO5 y oxígeno disuelto (Qi & Wu, 2023; Ahmed et al., 2021).

Análisis de las matrices de confusión

Las matrices de confusión revelaron que las clases “regular” y “mala” fueron las mejor clasificadas, mientras que las clases “mala” y “buena” presentaron algunas confusiones, probablemente debido a la similitud en sus perfiles fisicoquímicos. Este comportamiento sugiere que algunas zonas con contaminación incipiente podrían estar oscilando entre dos categorías, lo que justificaría un sistema de clasificación probabilística más que determinista (Zhang et al., 2022).

Incidencia del preprocesamiento

La limpieza de valores extremos (como un pH de 8.004 o DBO5 de 12.765) y la normalización de clases fueron determinantes para lograr una clasificación confiable. Sin este paso, los modelos sufrían de sobreajuste y degradación del desempeño, como ha sido documentado en trabajos previos (Liu et al., 2023; Gholami et al., 2020).

Además, la transformación min-max permitió estabilizar el entrenamiento de la red neuronal y facilitar la convergencia del modelo. Esta etapa también redujo la influencia de variables con escalas elevadas, como los coliformes fecales, que tendían a dominar la segmentación si no se ajustaban.

Interpretación ambiental de los resultados

El análisis evidencia que los parámetros fisicoquímicos monitoreados son suficientes para predecir con alta precisión la categoría de calidad del agua, lo que valida el enfoque adoptado en el proyecto de vinculación. La incorporación de modelos como Random Forest y MLP permite transformar estos datos en herramientas de alerta temprana, facilitando la toma de decisiones por parte de autoridades ambientales y comunidades locales.

Además, los resultados respaldan la hipótesis del estudio: el preprocesamiento adecuado mejora significativamente la precisión de los modelos de clasificación de calidad del agua, y puede ser integrado a sistemas inteligentes de monitoreo ambiental en zonas rurales de difícil acceso.

CONCLUSIONES

1. El preprocesamiento de datos fue esencial para garantizar la calidad del análisis. La eliminación de valores atípicos extremos ($\text{pH} > 14$, $\text{DBO5} > 100$), la estandarización de etiquetas y la normalización de variables permitieron mejorar el rendimiento de los algoritmos aplicados y asegurar la confiabilidad del proceso de clasificación. Estos pasos coincidieron con las mejores prácticas descritas en la literatura científica (Liu et al., 2023; Kotsiantis et al., 2006).
2. El modelo de Red Neuronal Artificial mostró un rendimiento ligeramente superior. Con un accuracy de 87.64 %, un F1-score de 0.8675 y una mejor capacidad de generalización frente a la complejidad de los datos, el modelo neuronal superó marginalmente a Random Forest, que obtuvo un accuracy de 87.48 % y un F1-score de 0.8613.
3. Los parámetros fisicoquímicos más relevantes para la clasificación fueron DBO5, coliformes fecales, oxígeno disuelto y nitratos. Estos parámetros fueron identificados por el modelo Random Forest como los más importantes en la predicción de la calidad del agua, lo cual se alinea con la normativa internacional y estudios previos sobre cuerpos hídricos tropicales (Gholami et al., 2020; WHO, 2017).
4. La clase “regular” fue la predominante y presentó mayor variabilidad. De las 2.082 muestras limpias, 1450 fueron clasificadas como “regular”, lo cual puede indicar zonas de transición entre estados de conservación y deterioro. Esta alta proporción sugiere la necesidad de monitoreo permanente para evitar que estas áreas evolucionen hacia condiciones “malas”.

El uso de mapas temáticos permite visualizar espacialmente los focos de deterioro ambiental. El mapa generado con coordenadas UTM permitió identificar que los puntos “mala” se concentran en zonas bajas, posiblemente afectadas por actividades humanas como asentamientos, vertidos domésticos o minería artesanal.

Referencias

1. Ahmed, M. U., Rauf, H. T., Lali, M. I. U., & Khan, M. A. (2021). A machine learning framework for water quality prediction using deep learning models. *IEEE Access*, 9, 79304–79317. <https://doi.org/10.1109/ACCESS.2021.3084200>
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
4. Gholami, A., Banejad, H., & Nazari, S. (2020). Water quality classification using machine learning models: A case study. *Environmental Monitoring and Assessment*, 192(4), 1–17. <https://doi.org/10.1007/s10661-020-08216-4>
5. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182. <http://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf>
6. Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, 50(2), 105–115. <https://doi.org/10.1016/j.artmed.2010.05.002>
7. Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Data preprocessing for supervised learning. *International Journal of Computer Science*, 1(2), 111–117. <https://arxiv.org/abs/2007.06044>
8. Liu, H., et al. (2023). Integrated approach to assess river water quality using data preprocessing and ensemble machine learning. *Environmental Research*, 224, 115504. <https://doi.org/10.1016/j.envres.2023.115504>
9. MAATE. (2022). Informe de calidad ambiental hídrica 2022. Ministerio del Ambiente, Agua y Transición Ecológica del Ecuador. <https://ambiente.gob.ec/informes-de-calidad-ambiental-hidrica/>
10. Organización Mundial de la Salud (OMS). (2017). *Directrices para la calidad del agua potable* (4.^a ed.). OMS. <https://www.who.int/publications/i/item/9789241549950>

11. Pourghasemi, H. R., & Gayen, A. (2020). Water quality index prediction using ensemble machine learning algorithms. *Science of the Total Environment*, 731, 139211. <https://doi.org/10.1016/j.scitotenv.2020.139211>
12. Qi, Y., & Wu, J. (2023). Assessment of classification models for monitoring river pollution. *Water Research*, 239, 120011. <https://doi.org/10.1016/j.watres.2023.120011>
13. Rodríguez-Pérez, R., Muñoz-Tapia, M., & Andrade, M. (2020). Diagnóstico de la calidad de agua superficial de los ríos ecuatorianos. *Revista Politécnica*, 45(1), 77–89. https://revistapolitecnica.epn.edu.ec/index.php/revista_politecnica/article/view/1221
14. WHO. (2017). *Guidelines for drinking-water quality* (4th ed.). World Health Organization. <https://www.who.int/publications/i/item/9789241549950>
15. Zhang, Y., Wang, Y., & Zhuang, X. (2022). Comparative analysis of ML algorithms in classifying water quality. *Environmental Pollution*, 297, 118741. <https://doi.org/10.1016/j.envpol.2021.118741>
16. Adhikari, B. R., & Shrestha, N. (2020). Assessment of surface water quality using water quality index and multivariate statistical techniques in Bagmati River, Nepal. *Environmental Challenges*, 1, 100003. <https://doi.org/10.1016/j.envc.2020.100003>
17. Bhardwaj, S. K., & Poonia, R. C. (2021). A machine learning-based approach for predicting water quality index using ensemble learning. *Environmental Science and Pollution Research*, 28(19), 23891–23903. <https://doi.org/10.1007/s11356-020-11623-3>
18. Ren, X., Liu, X., & Lin, M. (2022). A hybrid machine learning model for water quality classification based on SVM and PCA. *Environmental Monitoring and Assessment*, 194, 799. <https://doi.org/10.1007/s10661-022-10337-y>
19. Chen, K., Lu, W., Wu, W., & Zhou, L. (2023). A spatiotemporal analysis of river water quality using deep learning and satellite data. *Science of the Total Environment*, 858, 159847. <https://doi.org/10.1016/j.scitotenv.2022.159847>
20. Hossain, M. S., Chowdhury, M. A. M., & Rahman, M. M. (2019). Water quality classification using artificial neural network and support vector machine: A case study of the Buriganga River, Bangladesh. *Sustainable Water Resources Management*, 5, 1347–1355. <https://doi.org/10.1007/s40899-019-00315-3>.

© 2025 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).