

article_ronny.docx

 My Files My Files Universidad Peruana Union

Detalles del documento

Identificador de la entrega

trn:oid:::29566:431742018

Fecha de entrega

19 feb 2025, 10:27 a.m. GMT-5

Fecha de descarga

20 feb 2025, 10:26 a.m. GMT-5

Nombre de archivo

article_ronny.docx

Tamaño de archivo

2.4 MB

17 Páginas

6,995 Palabras

39,710 Caracteres

15% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado
- Coincidencias menores (menos de 8 palabras)

Exclusiones

- N.º de fuentes excluidas

Fuentes principales

- 10%  Fuentes de Internet
- 3%  Publicaciones
- 9%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Fuentes principales

- 10% Fuentes de Internet
- 3% Publicaciones
- 9% Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	Internet	www.researchgate.net	<1%
2	Internet	worldwidescience.org	<1%
3	Internet	dspace.esPOCH.edu.ec	<1%
4	Internet	tesis.ucsm.edu.pe	<1%
5	Internet	www.coursehero.com	<1%
6	Trabajos entregados	Universidad Cesar Vallejo on 2023-12-14	<1%
7	Internet	repository.upb.edu.co	<1%
8	Trabajos entregados	Consorcio CIXUG on 2023-07-27	<1%
9	Internet	laccEI.org	<1%
10	Trabajos entregados	ITESM: Instituto Tecnológico y de Estudios Superiores de Monterrey on 2025-01-30	<1%
11	Trabajos entregados	Universidad Carlos III de Madrid - EUR on 2024-06-05	<1%

12	Trabajos entregados	Universidad San Francisco de Quito on 2024-04-22	<1%
13	Internet	repositorio.unj.edu.pe	<1%
14	Publicación	Christian Xavier Meregildo Collave, Robert Jefferson Lázaro Bacilio, Adolfo Enriqu...	<1%
15	Trabajos entregados	Universidad Andina del Cusco on 2020-03-02	<1%
16	Trabajos entregados	Universidad Cesar Vallejo on 2019-01-17	<1%
17	Internet	editoraartemis.com.br	<1%
18	Internet	latam.hach.com	<1%
19	Internet	www.theinsightpartners.com	<1%
20	Trabajos entregados	Institución Universitaria Digital de Antioquia on 2024-12-09	<1%
21	Trabajos entregados	Centro de Investigación y Docencia Económicas AC on 2019-06-12	<1%
22	Trabajos entregados	Universidad Internacional de la Rioja on 2023-05-10	<1%
23	Trabajos entregados	Universidad Internacional de la Rioja on 2024-11-25	<1%
24	Internet	revistatyc.org.mx	<1%
25	Publicación	Pablo Andrés Buestán Andrade, Pedro Esteban Carrión Zamora, Anthony Eduardo...	<1%

26	Trabajos entregados	Universidad de Guadalajara on 2024-07-19	<1%
27	Internet	repositorio.espe.edu.ec	<1%
28	Trabajos entregados	Instituto Tecnológico de Costa Rica on 2024-04-24	<1%
29	Trabajos entregados	Universidad Arturo Prat on 2025-02-01	<1%
30	Trabajos entregados	Universidad Autónoma de Ciudad Juárez on 2025-02-18	<1%
31	Trabajos entregados	Universidad Internacional de la Rioja on 2025-01-27	<1%
32	Internet	ia802804.us.archive.org	<1%
33	Internet	pesquisa.bvsalud.org	<1%
34	Internet	researchoutput.csu.edu.au	<1%
35	Internet	wikileaks.org	<1%
36	Internet	www.ayto.zaragoza.es	<1%
37	Internet	www.businessmarketinsights.com	<1%
38	Internet	www.equipoweb.com.ar	<1%
39	Internet	www.mrt.com	<1%

40	Trabajos entregados	Politécnico Colombiano Jaime Isaza Cadavid on 2016-04-05	<1%
41	Publicación	Tony Marks-Block, Frank K. Lake, Rebecca Bliege Bird, Lisa M. Curran. "Revitalized...	<1%
42	Trabajos entregados	Universidad Continental on 2023-02-17	<1%
43	Trabajos entregados	Universidad Francisco de Vitoria on 2024-04-25	<1%
44	Trabajos entregados	Universidad Internacional de la Rioja on 2023-05-10	<1%
45	Internet	dspace.unitru.edu.pe	<1%
46	Internet	es.scribd.com	<1%
47	Internet	mundoagropecuario.com	<1%
48	Internet	www.jove.com	<1%
49	Internet	www.przetargi.info	<1%
50	Internet	www.puntoempresas.cl	<1%
51	Internet	www.slideshare.net	<1%
52	Publicación	"Discrete choice models applied to revenue management systems for air transpo...	<1%
53	Publicación	García, . "Long term performance", Dam Maintenance and Rehabilitation II, 2010.	<1%

54	Trabajos entregados	ITESM: Instituto Tecnológico y de Estudios Superiores de Monterrey on 2024-07-22	<1%
55	Trabajos entregados	Infile on 2024-02-19	<1%
56	Trabajos entregados	Universidad Carlos III de Madrid - EUR on 2023-09-03	<1%
57	Trabajos entregados	Universidad De Cuenca on 2023-09-14	<1%
58	Trabajos entregados	Universidad Rey Juan Carlos on 2024-06-07	<1%
59	Internet	aguamarket.com	<1%
60	Internet	bdigital.zamorano.edu	<1%
61	Internet	digibuo.uniovi.es	<1%
62	Internet	futur.upc.edu	<1%
63	Internet	rstudio-pubs-static.s3.amazonaws.com	<1%
64	Internet	scienceducation.galeon.com	<1%
65	Trabajos entregados	uncedu on 2024-02-29	<1%
66	Trabajos entregados	upeu on 2024-10-04	<1%
67	Internet	www.dspace.uce.edu.ec	<1%

68

Internet

www.jourlib.org

<1%

Predicción de la dosificación de coagulantes mediante random forest en la PTAP Miguel de la Cuba Ibarra

Ronny Iván Gonzales Medina, Rosa María Morán Silva y Javier Linkolk López Gonzales

Resumen

El agua potable es un recurso esencial para la humanidad con un impacto ambiental significativo. Por lo tanto, su tratamiento debe ser una prioridad. La predicción de la dosis de coagulantes mediante modelos y técnicas de Machine Learning permite mejorar la eficiencia y reducir los costos en el tratamiento del agua. Este enfoque debe ser implementado y evaluado, especialmente en países con recursos limitados. En este estudio, se aplicó el modelo de Random Forest para optimizar la predicción de la dosificación de coagulantes en la PTAP Miguel de la Cuba Ibarra. Los resultados obtenidos muestran que el modelo RF es adecuado para predecir niveles de turbidez de agua decantada y con la plataforma digital permite generar diferentes curvas de dosificación para la toma de decisiones en el menor tiempo posible.

Introducción

Las plantas de tratamiento de agua potable desempeñan un papel crucial en la sociedad, ya que proporcionan el elemento líquido indispensable para la supervivencia humana (Mishra, 2023). La calidad del agua se evalúa mediante diversos descriptores físicos como la turbidez, los sólidos totales, sedimentables, filtrables y disueltos, así como, el color, olor y temperatura (Wu et al., 2021). La predicción de los cambios en la calidad del agua permite identificar algún desbalance fisicoquímico del agua y tomar medidas de protección o remediación, con el objetivo de garantizar la calidad de la misma (Georgescu et al., 2023; Mănoiu y Crăciun, 2021). Las plantas de tratamiento son responsables de tratar el agua cruda y transformar el agua apta para el consumo bajo estándares de calidad según las entidades sanitarias de cada país (Katrivesis et al., 2019) (Wu, 2020). El proceso de transformación inicia con coagulación-floculación, proceso sedimentación y filtración, donde la etapa de coagulación es la más importante, ya que es el primer paso en la eliminación de partículas en suspensión, se agregan productos químicos, conocidos como coagulantes y floculantes (Sheng et al., 2023). Los coagulantes y floculantes más comunes utilizados en el tratamiento de agua son el sulfato de aluminio, policloruro de aluminio y polímero no iónico, que se agregan al agua en pequeñas cantidades, después de la coagulación, el agua se dirige a un tanque de sedimentación, donde los flóculos más grandes y pesados se asientan en el fondo del tanque. El agua sobrenadante, que contiene menos partículas en suspensión es llamada agua decantada (Salem et al., 2021).

El control de la reacción de coagulación es complejo y, a menudo, inalcanzable mediante métodos convencionales. La dosificación del coagulante generalmente se establece a través de la prueba de jarras, la cual demanda un tiempo prolongado de experimentación en una planta de tratamiento de agua en el campo (Yu et al., 2000). La modelización puede emplearse para superar estas limitaciones, lo que permite una predicción más precisa de la dosis de coagulantes. Esto mejora la eficiencia y reduce los costos en el tratamiento del agua (Kim & Parnichkun, 2017). Aunque se han desarrollado numerosos modelos y técnicas para predecir la dosis óptima de coagulantes, la mayoría se fundamenta en métodos estadísticos tradicionales y modelos lineales (Achite et al., 2023). Lo que genera limitaciones en la complejidad y variabilidad de los datos utilizados, siendo una de sus principales debilidades el sobreajuste, que puede

conducir a una reducción en la generalización del modelo ([Ahmed et al., 2023](#)). Uno de los enfoques con mayor alcance y aun expansión es el modelo Random Forest que brinda una mejor precisión de clasificación requiere menos tiempo de entrenamiento que Decision Tree y Support Vector Machine, por lo que es rápido. Tiene una fuerte capacidad de aprendizaje adaptativo y capacidad de mapeo no lineal, muy útil en grandes conjuntos de datos y también evita el problema de sobreajuste al manejar los ruidos presentes en los conjuntos de datos ([Bhattacharjee, A., Murugan et al., 2022](#); [Liu et al., 2012](#); [Parmar et al., 2019](#); [Sun et al., 2020](#)).

En este trabajo, se propone el uso de Random Forest, para predecir la dosis óptima del coagulante y floculante comúnmente utilizados en la coagulación del agua: sulfato de aluminio, policloruro de aluminio (PAC) y polímero, se recopilaron datos de calidad del agua y dosis de coagulantes y se utilizaron para entrenar y validar el modelo de Random Forest. La predicción de las dosis de coagulación también reducirá el impacto ambiental y mejorará la eficiencia de la planta de tratamiento. Al regular la cantidad de coagulante agregado al agua cruda en tiempo real, es posible reducir la cantidad de coagulante desperdiciado y optimizar el rendimiento de la planta de tratamiento de agua, al tiempo que se reduce la posibilidad de errores humanos en el proceso de toma de decisiones.

Trabajos Relacionados

La búsqueda de una mayor precisión, eficiencia conlleva en la actualidad al empleo de modelos de automatización (Machine Learning) tales como, redes neuronales (RN), máquinas de soporte vectorial (SVM) y Random Forest (RF). El uso de técnicas de aprendizaje automático puede mejorar la precisión y eficiencia en la predicción de la dosis óptima de coagulantes en la clarificación del agua. ([Bressane, et al., 2023](#)). Tal como, lo refiere [Zhang et al. \(2013\)](#) quienes investigaron dos métodos de aprendizaje automático, las máquinas de soporte vectorial (SVM) y el método de los vecinos más cercanos (KNN), para pronosticar las dosis de coagulante en diversas plantas de tratamiento de agua. Los resultados mostraron que las técnicas de regresión basadas en SVM demostraron una capacidad de pronóstico adecuada en plantas de tratamiento de agua grandes y medianas, mientras que KNN fue más efectivo para plantas de tratamiento de agua pequeñas. En otro estudio, [Gomes et al.\(2015\)](#) exploraron la coagulación del agua empleando dos coagulantes, sulfato de aluminio (AS) y policloruro de aluminio (PAC), y evaluaron distintos modelos de redes neuronales para prever las dosis de AS y PAC. Los resultados mostraron coeficientes de determinación de 0.95 y 0.91, respectivamente, lo que indicó una precisa capacidad de predicción.

Además, [Baouab Cherif \(2018\)](#) propusieron un enfoque para superar las limitaciones de las pruebas convencionales en plantas de tratamiento de agua, mediante la construcción de modelos basados en la turbidez, conductividad y pH del agua cruda y tratada. Desarrollaron dos modelos: uno aplicable cuando la turbidez era inferior a 45.5 NTU y el otro en caso contrario. Ambos modelos mostraron coeficientes de correlación superiores a 0.8 y errores medios absolutos de 5.47 g/m3 y 5.69 g/m3, respectivamente, en la dosificación.

[Kote y Wadkar \(2019\)](#) investigaron la aplicación de la FFNN, la RBFNN, la GRNN y la red neuronal feed-forward en cascada (CFNN) para las dosis de cloro y coagulante en una planta de tratamiento de agua de la India. Estos modelos se construyeron alterando las variables de entrada, las funciones de entrenamiento, los nodos ocultos, las épocas y los factores de propagación para mejorar sus rendimientos. Tras evaluar los indicadores de rendimiento, RBFNN con $R=0,999$ resultó ser el modelo de dosificación de cloro con mejor rendimiento. Del mismo modo, utilizando el diseño de red (2-40-1), el modelo CFNN con función de entrenamiento de regularización bayesiana ($R=0,947$) generó excelentes estimaciones para la dosis de coagulante. [Asmel et al.\(2021\)](#) sugirieron un método alternativo a las pruebas de jarras para determinar las dosis óptimas de sulfato de aluminio en plantas de tratamiento de agua. Utilizaron

una Red Neural de Base Radial (RBNN) para anticipar la turbidez y el pH del efluente en el tratamiento del agua. Los resultados revelaron que la RBNN permitió la predicción en tiempo real de la dosis óptima, evitando la repetición de pruebas de jarras.

Heddam (2021) aplicó los modelos de árbol extremadamente aleatorio (ERT), bosque aleatorio (RF) y regresión lineal múltiple (MLR). Los modelos se construyeron utilizando una variedad de factores de calidad del agua como entradas, incluyendo temperatura del agua, pH, turbidez, oxígeno disuelto y conductividad eléctrica. Para evaluar la precisión de los modelos se utilizaron las métricas R, error medio absoluto (MAE), error cuadrático medio (RMSE) y eficiencia Nash-Sutcliffe (NSE). Según los resultados, el modelo ERT fue el más resistente de todos los examinados. Tanto el modelo ERT como el RF tuvieron una buena precisión de entrenamiento y validación; sin embargo, el modelo MLR tuvo la menor precisión de entrenamiento y validación. Durante la fase de validación, el modelo ERT obtuvo los valores R y NSE más altos, 0,899 y 0,790, respectivamente. Estos valores fueron superiores a los del modelo RF ($R=0,851$ y $NSE=0,708$) y significativamente superiores a los del modelo MLR. Esto demostró que el modelo ERT presentaba el mejor rendimiento para predecir la dosis de coagulante en las plantas de tratamiento de agua potable.

Narges et al.(2021) aplicaron un sistema de inferencia neurofuzzy adaptativo (ANFIS) utilizando el enfoque de clustering sustractivo para calcular la dosis del coagulante óptimo en instalaciones de tratamiento de aguas. La técnica de agrupamiento tendió a reducir el número de reglas y vínculos, mejorando al mismo tiempo el reconocimiento y la responsabilidad del modelo. Los datos de entrada incluían la turbidez del agua, el pH, la temperatura, la alcalinidad y la conductividad eléctrica. Los resultados informaron de que el modelo ANFIS se asoció con R de 0,85 y 0,84 y RMSE de 1,32 y 1,83 para modelar las dosis de coagulante de alumbre y cloruro de polialuminio, respectivamente. Esto confirmó que ANFIS era un enfoque excelente para determinar la mejor dosis de coagulación en la planta de tratamiento de agua.

Tahraoui et al.(2021) predijeron la dosis de coagulante utilizando cuatro modelos matemáticos: red neuronal artificial (RNA), ANFIS, máquina de soporte vectorial (SVM) y metodología de superficie de respuesta (RSM). Aunque el rendimiento del modelo RNA fue ligeramente superior al de los demás modelos, los resultados mostraron que todos los modelos se ajustaban con precisión a los datos experimentales. Los resultados de los modelos RNA y SVM fueron idénticos. Sin embargo, en la fase de validación, el coeficiente de correlación del modelo RNA era muy alto, y los indicadores estadísticos eran bastante bajos en comparación con el modelo SVM. A pesar de ello, el modelo SVM era preferible al modelo RNA porque estaba asociado a menos parámetros. El modelo RSM demostró su eficacia por sus elevados coeficientes con el número de parámetros.

Wadkar et al.(2021) implementaron el modelo CFNN utilizando los algoritmos de entrenamiento Levenberg-Marquardt y Regularización Bayesiana para estimar la dosis de coagulante. Los nodos ocultos variaron de 15 a 60 durante el desarrollo de estos modelos, y R estuvo entre 0,914 y 0,947. El modelo CFNN con 40 nodos ocultos y el algoritmo de entrenamiento de Regularización Bayesiana produjo los mejores resultados, con $R=0,945$ y 0,947 para las fases de entrenamiento y prueba, respectivamente. Por otro lado, Jayaweera y Aziz (2018) predijeron la dosis óptima utilizando modelos como el perceptrón multicapa (MLP), así como la máquina de aprendizaje extremo (ELM) acoplada con la red neuronal feed-forward (ELM-FFNN) y la red neuronal de función de base radial (ELM-RBF). Todas las redes neuronales obtuvieron buenos resultados, con un coeficiente de correlación (R) superior a 0,97. Con una mayor precisión de predicción, la red ELM-RBF superó al modelo MLP, lo que demuestra su capacidad para la predicción de dosis de coagulante en el tratamiento del agua.

Asimismo, Jayaweera Aziz (2022) desarrollaron un modelo para determinar eficientemente la dosis de coagulante, tomando en cuenta los cambios en los parámetros del agua cruda (pH, temperatura,

59 turbidez). Utilizaron los modelos GRNN y ELM-RBF para anticipar datos de prueba con valores R de 0.9737 y 0.9783, respectivamente. El modelo ELM-RBF demostró una mayor capacidad de generalización que el GRNN, convirtiéndose en el más adecuado para la predicción de la dosis de coagulante.

De igual manera, Wang et al.(2023) en China propusieron los modelos de red neuronal de Elman (ENN) y bosque aleatorio (RF) para predecir con precisión la dosis de coagulante requerida y la turbidez del efluente. Los resultados mostraron que el ENN predijo mejor la dosis de coagulante, los dos modelos presentaron mejor desempeño para estas dos tareas registraron MAPE del 2,38% y 6,45%. Los residuos de predicción para el mejor rendimiento del modelo ENN estuvieron en el rango de $-1,0$ a $1,0$ mg L $^{-1}$, y los residuales de las predicciones del mejor modelo de RF para la turbidez del efluente estuvieron en el rango de $-0,2$ a $0,1$ NTU. Es Así, que estos modelos estabilizaron los residuos de las turbiedades del efluente y el objetivo dentro de $0,2$ NTU, y redujeron el costo del coagulante utilizado en la planta de agua en un 10%.

7 Achite et al. (2023) presentó un modelo de predicción para tomar decisiones proactivas sobre las dosis de coagulante basadas en cambios en las características del agua cruda. El modelo eliminó la necesidad de emplear productos químicos costosos para realizar pruebas periódicas de jarras y permite una respuesta rápida a cambios abruptos en la calidad del agua. El modelo de pronóstico se desarrolló utilizando un bosque aleatorio (RF) independiente y un modelo de RF hibridado con optimización de algoritmo genético (GA), concretamente GA-RF, para simular la dosis de coagulante en la planta de tratamiento de aguas residuales de Sidi Yacoub en Argelia. Este estudio estableció que el modelo más eficiente para evaluar el proceso de coagulación utilizando cuatro métricas de evaluación; coeficiente de correlación (CC), índice disperso (SI), índice de concordancia de Willmott (WI) y error porcentual absoluto medio (MAPE). Según los hallazgos, el modelo GA-RF (CC = 0,975, SI = 0,150, WI = 0,986 y MAPE = 7,9), que tiene en cuenta los parámetros de entrada de producción de agua cruda, turbidez del agua, conductividad y materiales suspendidos, supera a los demás modelos. Sawalkar et al.(2024) en la India, desarrollaron cuatro modelos, referentes, al modelo A (turbidez), el modelo B (pH, alcalinidad, temperatura, turbidez), el modelo C (pH, alcalinidad, temperatura) y el modelo D (alcalinidad y turbidez). Estos modelos fueron entrenados utilizando ANN y RF. Estas técnicas mejoraron la eficiencia y eficacia del proceso de tratamiento del agua y, en última instancia, mejoran la calidad y seguridad del agua tratada. Los resultados indicaron que los valores RMSE de los modelos ANN son comparativamente más bajos que los de RF. Comparando entre los modelos A, B, C y D, el modelo B y el modelo D exhiben un mejor rendimiento, con valores RMSE más bajos.

52 En conclusión, los estudios presentados muestran la viabilidad y eficacia del uso de modelos de predicción basados en redes neuronales, bosques aleatorios y algoritmos genéticos para optimizar las dosis de coagulante en plantas de tratamiento de agua, evitando el gasto excesivo y repetitivo de pruebas de jarras y mejorando la eficiencia en el proceso de coagulación. Estos enfoques representan una solución innovadora y efectiva para el tratamiento del agua en distintas condiciones y contextos, contribuyendo al abastecimiento de agua potable de calidad a la población.

Materiales y Métodos

1 Los materiales utilizados en la prueba de jarras incluyeron los coagulantes policloruro de aluminio (PAC) y sulfato de aluminio, así como el floculante polímero no iónico. Se emplearon recipientes de 2 L para contener el agua cruda durante el proceso de coagulación y floculación. Para la medición de los parámetros fisicoquímicos del agua tratada, se utilizaron equipos. La turbidez fue determinada mediante un turbidímetro con lámpara de tungsteno TL2300, EPA, con un rango de 0 a 4000 NTU. El pH se midió con un medidor de pH/mV de una sola entrada para laboratorio HQ411D. La temperatura del agua se registró

con un termómetro digital. La alcalinidad se determinó mediante titulación con ácido sulfúrico (H_2SO_4) 0.02 N, utilizando anaranjado de metilo para la alcalinidad total. La dureza cálcica se midió mediante titulación complexométrica con EDTA 0.01 M, empleando negro de eriocromo T (NET) como indicador. La conductividad eléctrica y los sólidos disueltos totales (TDS) se determinaron utilizando el medidor de conductividad/TDS para laboratorio HQ430D con célula CDC401 de grafito con 4 polos.

Es un estudio de tipo aplicado, con el propósito de determinar, predecir y optimizar la dosificación del coagulante en el tratamiento de agua potable a través de modelos de machine learning. Donde se tomó de muestras de aguas residuales denominada prueba de jarras, se creó una base de datos para entrenar el modelo a través del Machine Learning Random Forest.

Prueba de jarras.

La prueba de jarras es una etapa importante en el proceso de tratamiento de aguas, y requiere especial seguimiento, que permite aproximar la dosis óptima de coagulación, este proceso a continuación es detallado. El procedimiento inicial para llevar a cabo una prueba de jarras es la preparación de coagulantes en solución teniéndose como base para soluciones madre una concentración 10% en peso del soluto tanto para el policloruro de aluminio y sulfato de aluminio para posteriormente diluirla a soluciones de 1%, mientras que la solución madre de polímero no iónico tiene una concentración de 0.05% p/v.

El agua cruda es almacenada en volumen unitario de 2L para su dosificación. La mezcla (coagulantes – agua cruda) tiene los siguientes procesos:

Agitación durante 15 segundos con velocidad de 300 rpm.

Reposo 10 minutos.

Agitación durante 30 segundos con una velocidad de 300 rpm.

Agitación durante 30 minutos con una velocidad de 30 rpm.

Reposo de 30 minutos.

Finalmente, se toma muestras de cada jarra (volumen unitario) para tomar los parámetros de turbidez, pH, temperatura, alcalinidad y dureza.

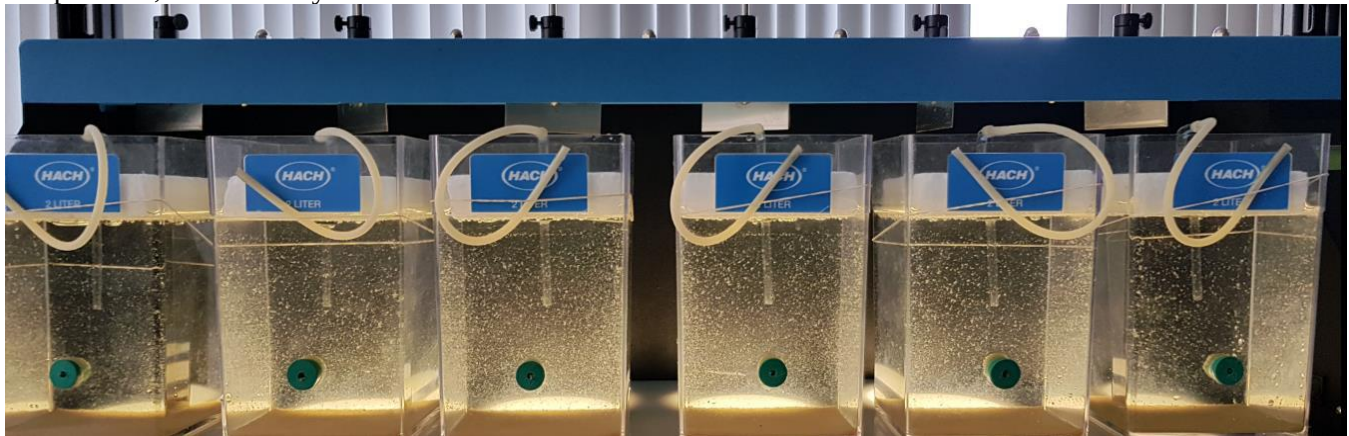


Figura 1. Pruebas de jarras en la planta Miguel de la Cuba Ibarra.

Durante el proceso se observó el tiempo de formación de floc, las natas que pueden llegar a formarse en la superficie del agua en cada jarra o el floc que se pega en los flotadores.

Modelo de Random Forest

La técnica de aprendizaje ensemble conocida como Random Forest (RF) combina la fuerza predictiva de muchos árboles de decisión con la adición de aleatorización para reducir el sobreajuste. RF genera una serie de árboles de decisión y utiliza bootstrapping para entrenar cada árbol en un subconjunto diferente de los datos de entrenamiento. La clasificación o predicción final se determina mediante una combinación de las

salidas individuales del árbol, que puede lograrse utilizando la mayoría de votos para tareas de clasificación o promediando para tareas de regresión.

Matemáticamente, representemos un conjunto de Random Forest como un conjunto de árboles de decisión $(T_1, T_2, \dots, T_n)$, donde cada árbol T_i , se entrena con una muestra bootstrap diferente del conjunto de datos de entrenamiento original. La clasificación de una nueva muestra X_{nuevo} mediante el conjunto Random Forest puede expresarse del siguiente modo: Para la clasificación:

$$\hat{y}_{RF}(X_{nuevo}) = \text{VotoMayoritario}(T_1(X_{nuevo}), T_2(X_{nuevo}), \dots, T_n(X_{nuevo})) \quad (1)$$

Para regresión:

$$\hat{y}_{RF}(X_{nuevo}) = \frac{1}{n} \sum_{i=1}^n T_i(X_{nuevo}) \quad (2)$$

Donde: $\hat{y}_{RF}(X_{nuevo})$ representa la predicción o clasificación final para la nueva muestra X_{nuevo} utilizando el conjunto Random Forest.

$T_i(X_{nuevo})$ representa la predicción o clasificación realizada por la i^{th} árbol de decisión en el conjunto para la nueva muestra. Voto mayoritario (.) calcula el voto mayoritario entre las predicciones de árboles individuales para tareas de clasificación, y n es el número total de árboles de decisión en el conjunto de RF. Después de todo, puede promediar o combinar las salidas de varios árboles, cada uno entrenado en un subconjunto separado de los datos, porque puede disminuir el sobreajuste. La resistencia del modelo y su capacidad de generalización mejoran con esta técnica de conjunto, lo que la convierte en una opción popular para diversas aplicaciones de aprendizaje automático.

Área de estudio y recolección de datos

La Planta de Tratamiento de Agua Potable (PTAP) Miguel de la Cuba Ibarra se encuentra en Arequipa, una ciudad ubicada en el sur de Perú, en una zona geográficamente diversa que abarca desde la costa hasta la cordillera de los Andes. Con un clima árido, con escasas precipitaciones anuales, lo que hace que el abastecimiento y tratamiento de agua sean especialmente críticos. La región se caracteriza por su terreno volcánico y suelos fértiles, siendo una zona de gran actividad agrícola. La cuenca del río Chili atraviesa la ciudad, proporcionando una fuente vital de agua que es tratada en plantas como la PTAP Miguel de la Cuba Ibarra para asegurar su potabilidad. La ubicación de la PTAP es estratégica, ya que se encuentra en una región donde la demanda de agua potable es alta debido al crecimiento poblacional y a las actividades industriales y agrícolas. El agua tratada en la planta llega a tener turbiedades de 500 NTU en los meses enero y febrero llegando hasta alcanzar un máximo de 3600 NTU. Los niveles bajos de turbidez se encuentran en los demás meses llegando a tener un rango 5 – 30 NTU.

Los datos de prueba de jarras fueron obtenidos de manera intermitente durante los periodos 2015-2022. Por cada registro de prueba de jarra se tuvo 6 experimentos. Los formatos pasaron por tres versiones, se completó los datos faltantes mediante promedios mensuales y técnicas de vecinos cercanos (KNN) en función al mes. Los datos contienen parámetros de turbidez (NTU), pH, Temperatura (°C), Conductividad (us/cm), solidos totales disueltos (mg/L), alcalinidad (mg/L), dureza cálcica (mg/L) para las propiedades de agua cruda, según estos parámetros se debe calcular la dosis de coagulantes para reducir la turbidez y llevarlas a valores menores a 2 NTU. Los coagulantes empleados para la dosificación son el policloruro de aluminio (mg/L) y sulfato de aluminio (mg/L), mientras que el floculante empleado es el polímero no iónico.

Resultados y Discusión

En el presente estudio, se utilizan modelos de Random Forest con dos combinaciones distintas de variables de entrada para estimar la dosis de coagulante y evaluar su precisión. El conjunto de datos empleado para construir el modelo incluye 426 pruebas de jarras, cuyas variables están detalladas en la Tabla 1. Estos datos se dividen en dos subconjuntos: el 70% para el entrenamiento del modelo y el 30% restante se emplea para la evaluación de su rendimiento.

Table 1. Resumen del Modelo Lineal.

Término	Estimado	Error Estándar	Estadístico t	Valor p	Significancia
(Intercept)	0.332187758	1.4799574997	0.22445763	8.224273e-01	
Tb_NTU	0.010452824	0.0006498164	16.08581011	1.992890e-54	***
pH	-0.016291807	0.1979451618	-0.08230465	9.344138e-01	
Temp_C	0.066841877	0.0123917051	5.39408226	7.819835e-08	***
Alc_CO3Ca	-0.001142291	0.0084848901	-0.13462646	8.929226e-01	
Ca2_mg. L	-0.000183768	0.0082677212	-0.02222717	9.822693e-01	
CND_us.cm	0.002821732	0.0007957449	3.54602651	4.013133e-04	***
PAC	-0.032675677	0.0032014127	-10.20664332	8.390891e-24	***
Sulfato_Al2	-0.032644798	0.0031443919	-10.38191140	1.506147e-24	***
Polimero	-2.424058937	0.4590850471	-5.28019580	1.450551e-07	***

El modelo desarrollado para predecir la dosis de coagulante está diseñado para aplicarse a aguas crudas con una turbidez inferior a 150 NTU, debido a la limitada disponibilidad de datos en rangos de turbidez superiores. En la Figura 2 se presenta la distribución de los datos en función de los intervalos de turbidez, evidenciando que más del 50% de los datos se concentran en intervalos de turbidez menores a 50 NTU. Aunque el comportamiento de la dosificación tiende a seguir una relación lineal para turbideces superiores al umbral de 50 NTU, facilitando la predicción hasta niveles de turbidez de 150 NTU, la distribución se vuelve más compleja para valores superiores a 150 NTU. Esta complejidad introduce un sesgo significativo en las predicciones para niveles de turbidez elevados. Para abordar este sesgo y minimizar el impacto de las variables en la predicción, se ha llevado a cabo un análisis de correlación y un análisis de varianza de un factor.

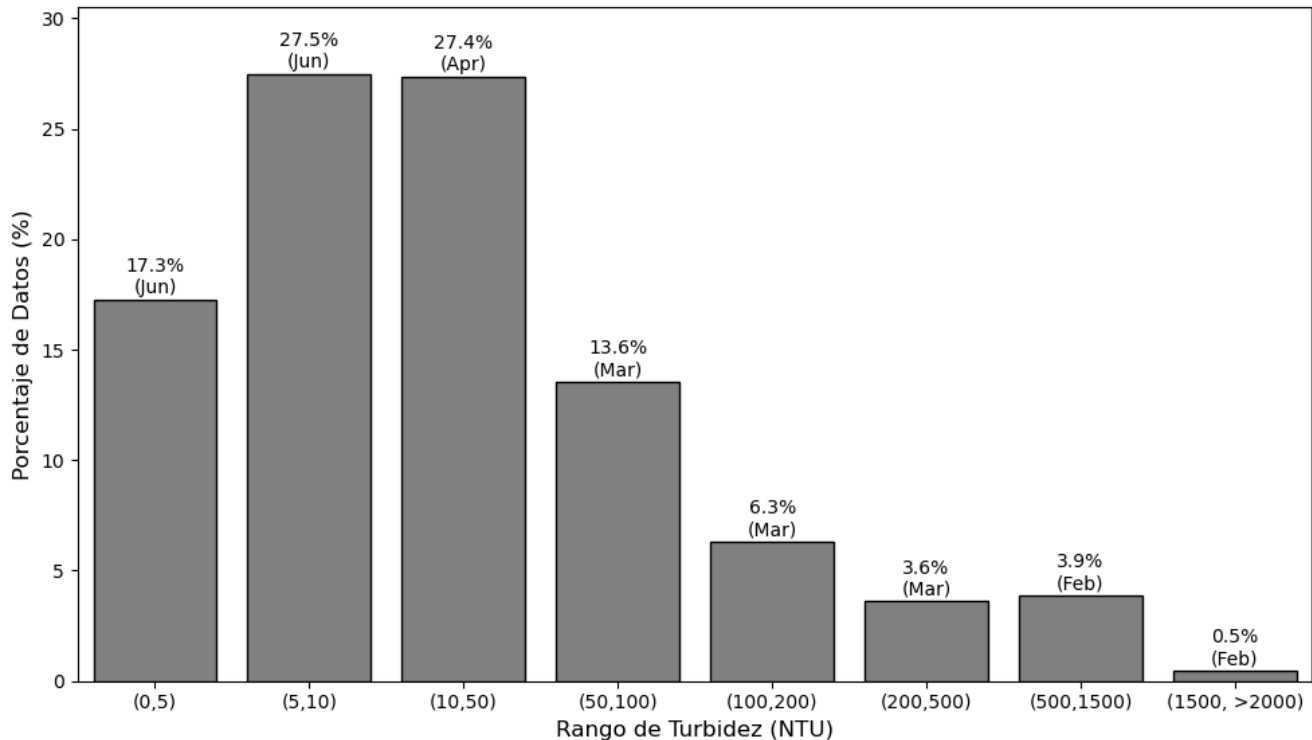


Figura 2. Distribución de turbidez del agua cruda.

La figura 3 muestra un análisis de correlación que permite observar las relaciones lineales entre varias variables predictoras. A partir de los resultados presentados, que aparecen como resultado de este gráfico, se pueden identificar algunas tendencias que ayudan a comprender qué variables son más relevantes para el modelo predictivo y cuáles pueden excluirse por su baja correlación o por ser redundantes.

En primer lugar, se observa que muchas de las variables están muy correlacionadas entre sí, lo que indica redundancias. Por ejemplo, existe una correlación positiva muy alta entre SDT (mg/L), CND (us/cm), Alc_CO3 y Ca2_mg/L ($R > 0,7$), de lo que se deduce que la variable captura la misma información en gran parte. Grandes correlaciones entre las variables independientes pueden generar problemas de multicolinealidad en este modelo, distorsionando los coeficientes y disminuyendo la interpretabilidad. En este caso, se debería suprimir una de estas variables para no perder capacidad predictiva y para que el modelo sea más eficiente, según lo expuesto por [Maier et al. \(2010\)](#). Este análisis de correlación subraya aún más cómo la selección adecuada de las variables para el modelo predictivo marca una gran diferencia. Al eliminar las variables correlacionadas de forma redundante, se evita el sobreajuste y, por tanto, se favorece el rendimiento del modelo, como se ha establecido en otros estudios que utilizan técnicas como las redes neuronales artificiales y los sistemas difusos adaptativos ([Kim & Parnichkun, 2017](#)). Una selección adecuada de las variables garantiza que el modelo se centre en las relaciones más importantes entre las variables predictoras y optimiza la precisión del modelo.

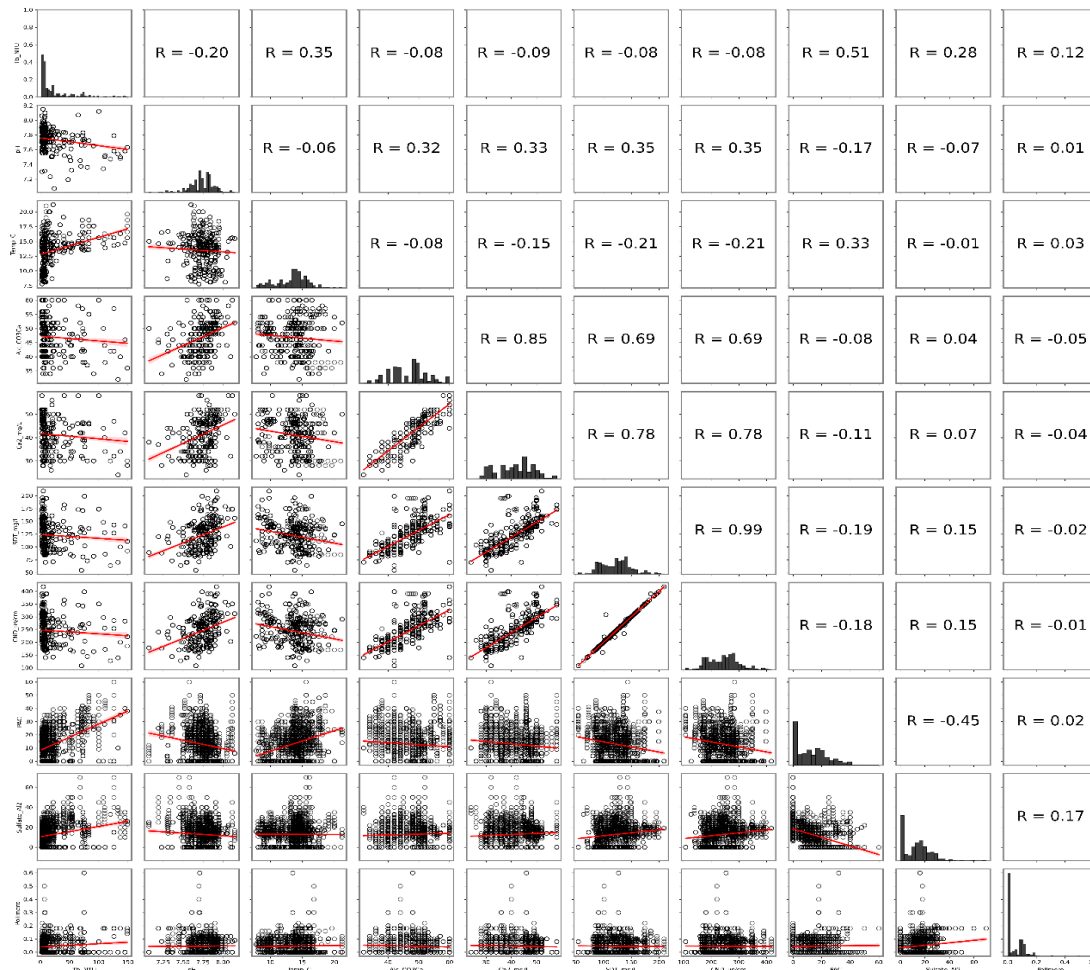


Figura 3. Matriz de correlación de predictores.

El análisis de regresión lineal múltiple proporciona una visión cuantitativa sobre la importancia de las variables predictoras en la turbidez del agua decantada. En primer lugar, el valor intercepto no es significativo (Valor $p = 0.82$), lo que indica que cuando todas las variables predictoras son cero, la turbidez predicha por el modelo no difiere significativamente de cero. Esto sugiere que el modelo está bien centrado y que las variaciones en la turbidez son realmente explicadas por los valores de las variables incluidas en el modelo. Sin embargo, el análisis de las variables predictoras individuales revela información crucial para ajustar y optimizar el modelo de predicción. La variable Tb_NTU, que representa la turbidez en otra fase o medida, muestra un coeficiente positivo y altamente significativo (Valor $p = 1.99e-54$). El valor estimado del coeficiente es 0.01045, lo que indica que un aumento de una unidad en esta variable genera un incremento predecible en la turbidez de aproximadamente 0.0104 unidades. Esto refuerza la idea de que la turbidez observada en etapas anteriores o en otras condiciones tiene una gran capacidad predictiva en el modelo. Este resultado es coherente con estudios previos que destacan la relevancia de la turbidez como variable clave en sistemas de predicción de calidad del agua mediante modelos de aprendizaje automático y estadísticas multivariadas (Kim & Parnichkun, 2017). Además, otras variables significativas como Temp_C (temperatura), CND_us.cm (conductividad), PAC (policloruro de aluminio) y Sulfato_Al2 (sulfato de aluminio) son igualmente importantes.

La temperatura (Temp_C) muestra un valor estimado positivo significativo de 0.0668 (Valor $p = 7.82e-08$), lo que implica que un aumento en la temperatura está asociado con un aumento en la turbidez. Este hallazgo es consistente con investigaciones que indican que las variaciones en la temperatura del agua afectan los procesos de coagulación y floculación, lo que a su vez influye en la turbidez del agua tratada (McKelvey et al., 2024)(Lin et al., 2023). Además, la conductividad (CND_us.cm) también es significativa (Valor $p = 4.01e-04$), lo que sugiere que los niveles de iones disueltos en el agua son un factor importante en la predicción de la turbidez. Esto se alinea con la literatura sobre el impacto de las propiedades fisicoquímicas del agua en los procesos de clarificación (Igwegbe et al., 2021). Por otro lado, las variables PAC y Sulfato_Al2 presentan coeficientes negativos significativos, con estimaciones de -0.0326 y valores p extremadamente bajos ($8.39e-24$ y $1.50e-24$, respectivamente). Esto indica que, a mayor dosificación de estos coagulantes, se espera una reducción en la turbidez, lo que es coherente con el papel que desempeñan estos compuestos en el tratamiento de agua. Estudios recientes en predicción de la turbidez también confirman que la correcta dosificación de coagulantes como el cloruro de polialuminio y el sulfato de aluminio es esencial para mantener la turbidez dentro de los límites aceptables en plantas de tratamiento de agua (Kennedy et al., 2015). La relación negativa y significativa de estos coagulantes con la turbidez es fundamental para optimizar los procesos de coagulación y floculación. Finalmente, la variable Polímero no iónico tiene una relación negativa altamente significativa con la turbidez (Valor $p = 1.45e-07$), lo que sugiere que el uso de polímeros en el tratamiento de agua también ayuda a reducir los niveles de turbidez. Los polímeros son comúnmente usados como agentes coagulantes para mejorar la sedimentación de partículas finas y reducir la turbidez, lo cual ha sido ampliamente documentado en la literatura (García Nieto et al., 2012). En conjunto, estos resultados destacan la importancia de seleccionar adecuadamente los parámetros de operación y dosificación de químicos en plantas de tratamiento de agua para garantizar niveles de turbidez adecuados y consistentes en el agua decantada.

El modelo de random forest fue construido utilizando el lenguaje de programación R con la librería randomForest, lo que permitió crear un modelo robusto para predecir la turbidez del agua decantada. En este análisis, se suprimieron las variables de sólidos totales disueltos (STD) y Ca dureza cálcica, debido a su alta colinealidad con otras variables, lo que podría haber afectado la precisión del modelo. La eliminación de estas variables evita la redundancia en los datos y ayuda a mejorar la eficiencia del modelo, al minimizar el problema de multicolinealidad, que ocurre cuando variables fuertemente correlacionadas afectan la capacidad del modelo para calcular con precisión la importancia de las predicciones. Esta decisión está en línea con la metodología propuesta por Maier et al. (2010), que sugiere eliminar variables redundantes para optimizar los modelos predictivos.

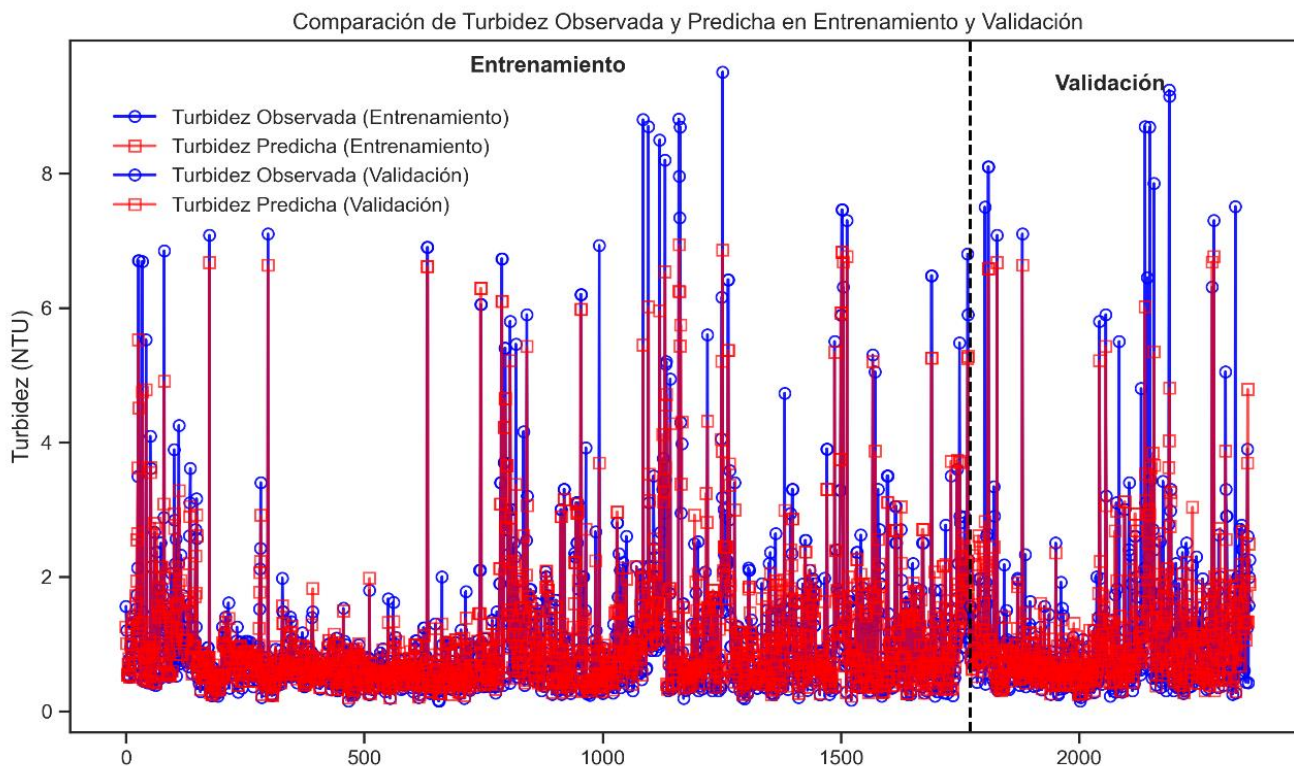


Figura 4. Evolución temporal de la turbidez tanto observada como predicha

La Figura 4 presenta la evolución temporal de la turbidez tanto observada como predicha en los conjuntos de entrenamiento y validación. En la parte de entrenamiento, el modelo sigue de cerca las fluctuaciones de los valores observados, lo que indica que el modelo de random forest captura adecuadamente la tendencia general de los picos de turbidez. Sin embargo, en el conjunto de validación, el modelo tiene mayores dificultades para predecir los picos extremos de turbidez. Esto podría deberse a la falta de representatividad de estos eventos en los datos de entrenamiento, lo que resalta la importancia de utilizar un conjunto de datos más grande y diverso para entrenar modelos predictivos, como se ha recomendado en estudios previos de predicción de la turbidez (Kennedy et al., 2015). Además, la eliminación de las variables correlacionadas, como Ca dureza cálcica y sólidos totales disueltos, fue clave para evitar la multicolinealidad y mejorar el rendimiento general del modelo, lo cual ha sido respaldado por la literatura en el campo del modelado predictivo de procesos de tratamiento de agua (McKelvey et al., 2024) (Lin et al., 2023). Los gráficos proporcionados presentan una visión clara de cómo el modelo de random forest predice la turbidez tanto en el conjunto de entrenamiento como en el de validación. En el primer gráfico, correspondiente al conjunto de entrenamiento, se puede observar una alta precisión del modelo con un R^2 de 0.92, lo que indica que el modelo explica el 92% de la variabilidad en los datos. El bajo valor de RMSE (error cuadrático medio) de 0.11 NTU y el MAPE (error absoluto porcentual medio) de 0.11 confirman la capacidad del modelo para realizar predicciones con un error mínimo en el conjunto de entrenamiento.

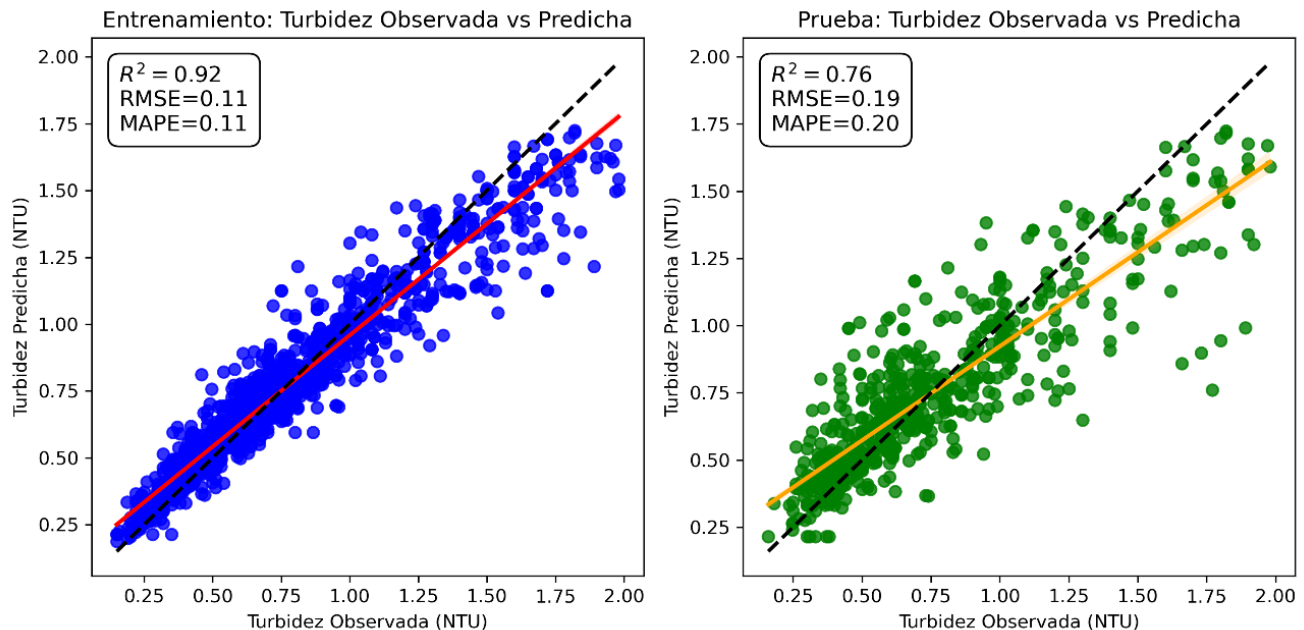


Figura 5. Predicciones del modelo

44

La línea roja en la figura 5 muestra que las predicciones del modelo están alineadas de manera efectiva con los valores observados, lo que indica que el modelo no está sobre ajustado. Sin embargo, en el segundo gráfico, correspondiente al conjunto de validación, el R^2 cae a 0.76, lo que sugiere una disminución en la precisión cuando el modelo se enfrenta a datos no utilizados durante el entrenamiento. Este comportamiento es común en los modelos de aprendizaje automático, donde el rendimiento suele disminuir en los conjuntos de prueba debido a la mayor complejidad de los datos no vistos. A pesar de esta disminución en el rendimiento, el RMSE de 0.19 y el MAPE de 0.20 siguen siendo relativamente bajos, lo que sugiere que el modelo sigue siendo efectivo para realizar predicciones de turbidez, aunque con un mayor margen de error comparado con los datos de entrenamiento.

El modelo desarrollado permitió la creación de una plataforma que sigue los pasos ilustrados en el gráfico 1. Esta plataforma facilita la evaluación de diversos escenarios de dosificación, dado que los datos correspondientes a la dosificación suelen ser periódicos.

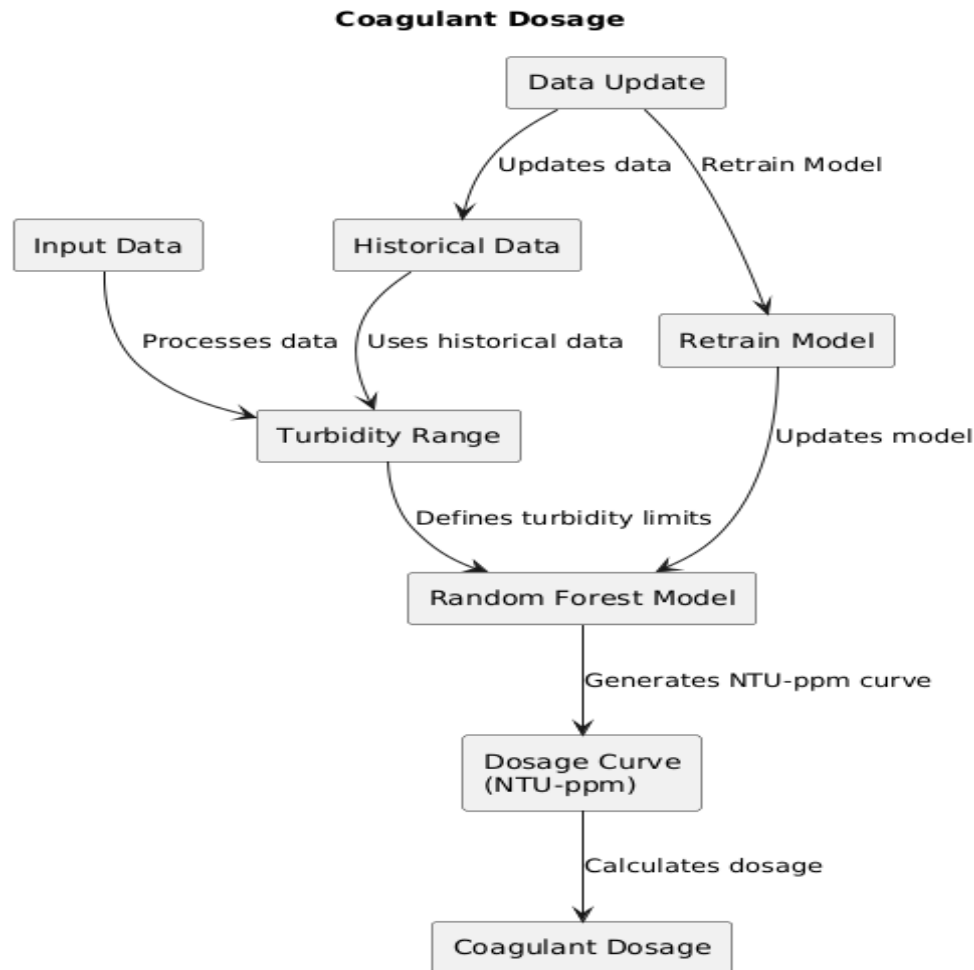


Gráfico 1: Pasos para la creación de la plataforma

En este contexto, se calculan las diferentes curvas de dosificación en función de la turbidez del agua cruda para los dos floculantes utilizados en la planta (PAC y sulfato de aluminio). Cada curva que se genera responde a la adición de diferentes concentraciones de polímero no iónico. El cálculo de las dosis de polímeros se realiza en función de los datos obtenidos a partir de la prueba de jarras, aplicando un margen del ± 50 % según la turbidez del agua ingresada. A partir de estos resultados, se determinan los datos correspondientes a la dosificación tanto del coagulante como del floculante. La figura 6 son herramientas clave para la toma de decisiones ante cambios drásticos en la turbidez durante la operación de la planta, asegurando una respuesta rápida y eficiente ante cualquier variación en las condiciones del agua.

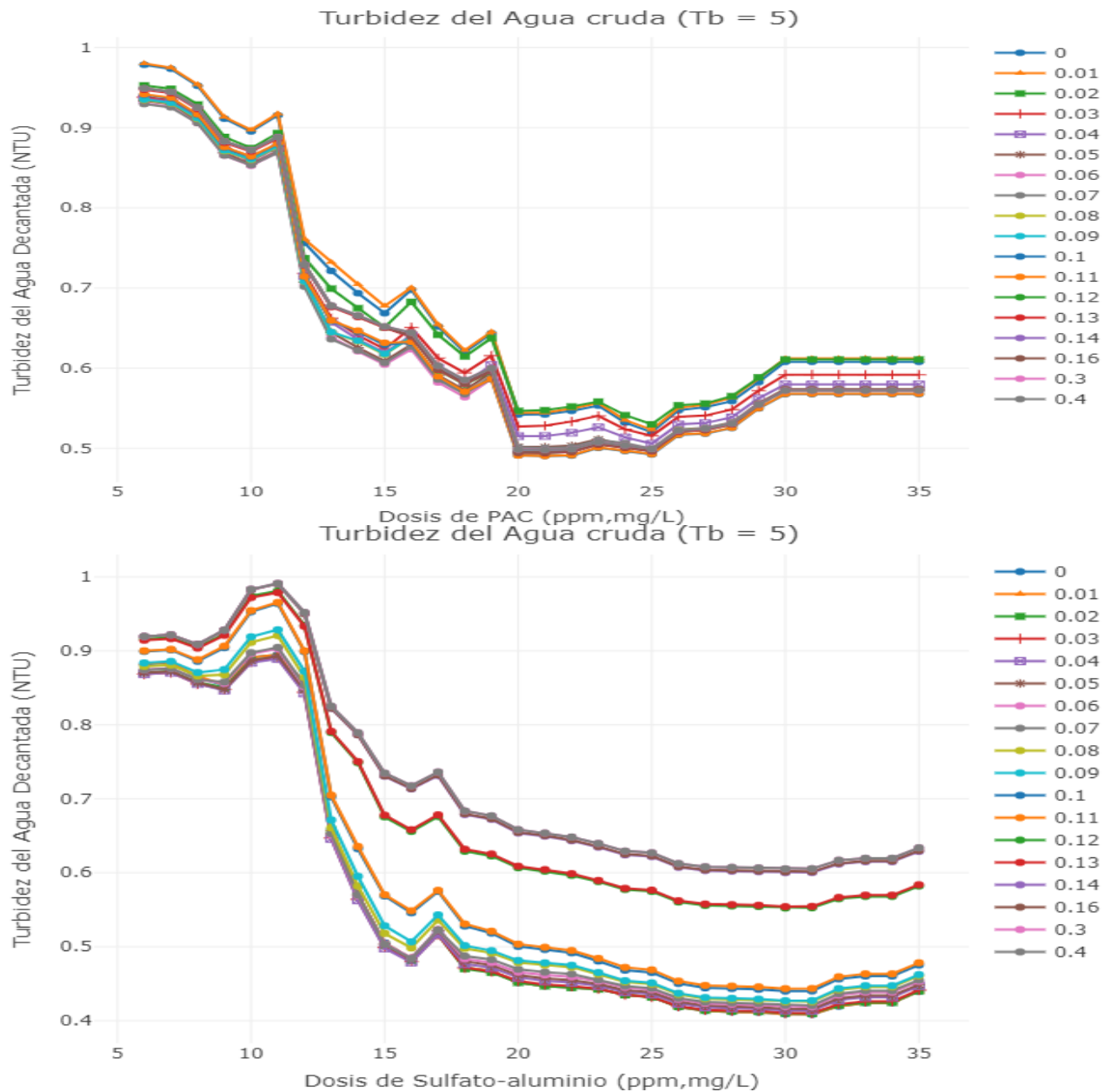


Figura 6: Curvas de Dosificación

Conclusión

El modelo de R-F predice la turbidez de agua decantada usando diferentes dosis de coagulantes. Forma diferente de evaluar las dosis de los coagulantes y concentraciones del polímero no aniónico. Se creó una plataforma digital para generar diferentes curvas de dosificación en función del agua cruda. El uso del polímero no aniónico debe ser cuidadosamente optimizado para evitar efectos negativos en la calidad del agua tratada.

Se identifica una dosis óptima de entre 10 y 20 ppm, donde la reducción de la turbidez es más efectiva, alcanzando valores cercanos a 0.5 NTU.

Referencias

- Achite, M., Samadianfard, S., Elshaboury, N. y Sharafi, M. . (2023). Modeling and optimization of coagulant dosage in water treatment plants using hybridized random forest model with genetic algorithm optimization. *Environ Dev Sustain* 25, 11189–11207 <https://doi.org/10.1007/s10668-022-02523-z>
- Ahmed, SF, Alam, MSB, Hassan, M., Rozbu, MR, Ishtiak, T., Rafa, N., ... y Gandomi, AH (2023). Técnicas de modelado de aprendizaje profundo: progreso actual, aplicaciones, ventajas y desafíos. *Artificial Intelligence Review* , 56 (11), 13521-13617. <https://doi.org/10.1007/s10462-023-10466-8>
- Asmel, NK, Al-Nima, RR, Mohammed, FI, Al Saadi, AM y Ganiyu, AA (2021). Pronóstico de la turbidez y el pH de los efluentes en una prueba de jarras utilizando una red neuronal de base radial. En *Towards a Sustainable Water Future: Proceedings of Oman's International Conference on Water Engineering and Management of Water Resources* (págs. 361-370). ICE Publishing <https://www.icevirtuallibrary.com/doi/epdf/10.1680/oicwe.65253.361>
- Bhattacharjee, A., Murugan, R., Soni, B. *et al.* Ada-GridRF: A Fast and Automated Adaptive Boost Based Grid Search Optimized Random Forest Ensemble model for Lung Cancer Detection. *Phys Eng Sci Med* 45, 981–994 (2022). <https://doi.org/10.1007/s13246-022-01150-2>
- Bhattacharya, A., Sahu, S., Telu, V., Duttagupta, S., Sarkar, S., Bhattacharya, J., ... y Ghosal, PS (2021). Neural Network and Random Forest-Based Analyses of the Performance of Community Drinking Water Arsenic Treatment Plants.. *Water* , 13 (24), 3507. <https://doi.org/10.3390/w13243507>
- Baouab, M. H., & Cherif, S. (2018). Prediction of the optimal dose of coagulant for various potable water treatment processes through artificial neural network. *Journal of Hydroinformatics*, 20(6), 1215-1226. <https://doi.org/10.2166/hidro.2018.014>
- Bressane, A., Goulart, A. P. G., Melo, C. P., Gomes, I. G., Loureiro, A. I. S., Negri, R. G., ... & Thomé, R. F. (2023). A non-hybrid data-driven fuzzy inference system for coagulant dosage in drinking water treatment plant: Machine-learning for accurate real-time prediction. *Water*, 15(6), 1126. <https://doi.org/10.3390/w15061126>
- García Nieto, P. J., Martínez Torres, J., Araújo Fernández, M., & Ordóñez Galán, C. (2012). Support machines and neural networks used to evaluate paper manufactured using Eucalyptus globulus. *Applied Mathematical Modelling*, 36(3), 6137-6145. <https://doi.org/10.1016/j.apm.2012.02.016>
- Georgescu, P-L., Moldovanu, S., Iticescu, C., Calmuc, M., Calmuc, V., Topa, C., Moraru, L. (2023) Assessing and forecasting water quality in the Danube River by using neural network approaches, *Science of The Total Environment* 879,162998, <https://doi.org/10.1016/j.scitotenv.2023.162998>.
- Gomes, LS, Souza, FAA, Pontes, RST, Fernandes Neto, TR, & Araújo, RAM (2015). Coagulant Dosage Determination in a Water Treatment Plant Using Dynamic Neural Network Models. *International Journal of Computational Intelligence and Applications*, 14(03), 1550013. <https://doi.org/10.1142/S1469026815500133>
- Hao, Z., AghaKouchak, A., Nakhjiri, N. *et al.* (2014)Global integrated drought monitoring and prediction system. *Sci Data* 1, 140001 <https://doi.org/10.1038/sdata.2014.1>

- Heddams, S. (2021). 24 - Extremely randomized tree: a new machines learning method for predicting coagulant dosage in drinking water treatment plant. *Water Engineering Modeling and Mathematic Tools*,. 475-489. <https://doi.org/10.1016/B978-0-12-820644-7.00013-X>
- Igwegbe, C. A., Onukwuli, O. D., & Ighalo, J. O. (2021). Bio-coagulation-flocculation of municipal solid waste leachate using Picralima nitida extract: RSM and ANN modelling. *Current Opinion in Green and Sustainable Chemistry*, 4, 100078. <https://doi.org/10.1016/j.crgsc.2021.100078>
- Jayaweera, C. D., & Aziz, N. (2018). Development and comparison of extreme learning machine and multi-layer perceptron neural network models for predicting optimum coagulant dosage for water treatment. *J. Phys.: Conf. Ser.* **1123** 012032 doi <https://doi.org/10.1088/1742-6596/1123/1/012032>
- Jayaweera, C. D., & Aziz, N. (2020). Prediction of treated water turbidity and its parametric behavior for the coagulation process of a water treatment plant using a joint extreme learning machine-genetic algorithm. *IOP Conference Series: Materials Science and Engineering*, 991(1). <https://doi.org/10.1088/1757-899X/991/1/012142>
- Jayaweera, C. D., & Aziz, N. (2022). An efficient neural network model for aiding the coagulation process of water treatment plants. *Environment, Development and Sustainability*, 24(1), 1069-1085. . <https://doi.org/10.1007/s10668-021-01483-0>
- Katrivesis, F. K., Karela, A. D., Papadakis, V. G. & Paraskeva, C. A. (2019) Revisiting of coagulation-flocculation processes in the production of potable water. *J. Water Process. Eng.* 27, 193–204, <https://doi.org/10.1016/j.jwpe.2018.12.007>
- Kennedy, M. J., Gandomi, A. H., & Miller, C. M. (2015). Coagulation modeling using artificial neural networks to predict both turbidity and DOM-PARAFAC component removal. *Journal of Environmental Chemical Engineering*, 3(4), 2829–2838. <https://doi.org/10.1016/j.jece.2015.10.010>
- Kim, C.M. & Parnichkun, M. (2017) Prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant using a hybrid model of *k*-means clustering and adaptive neuro-fuzzy inference system. *Appl Water Sci* 7, 3885–3902. <https://doi.org/10.1007/s13201-017-0541-5>
- Kote, A. S., & Wadkar, D. V. (2019). Modeling of chlorine and coagulant dose in a water treatment plant by artificial neural networks. *Eng. Technol. Appl. Sci. Res.*, 9(3), 4176-4181. <https://doi.org/10.48084/etasr.2725>
- Lin, S., Kim, J., Hua, C., Kang, S., & Park, M. H. (2023). Comparing artificial and deep neural network models for prediction of coagulant amount and settled water turbidity: Lessons learned from big data in water treatment operations. *Journal of Water Process Engineering*, 54, 103949.. <https://doi.org/10.1016/j.jwpe.2023.103949>
- Liu, Y., Wang, Y. & Zhang, J. (2012). New Machine Learning Algorithm: Random Forest. In: Liu, B., Ma, M., Chang, J. (eds) Information Computing and Applications. ICICA 2012. Lecture Notes in Computer Science, vol 7473. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34062-8_32
- Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental Modelling & Software*, 25(8), 891-909. <https://doi.org/10.1016/j.envsoft.2010.02.003>
- Mănoiu, V.-M., Crăciun, A.-I (2021) Danube river water quality trends: a qualitative review based on the open access web of science database Ecohydrol. *Hydrobiol.*, 21 (4), 613-628. <https://doi.org/10.1016/j.ecohyd.2021.08.002>
- McKelvey, S., Abassi, A., Nataraj, C., & Duran, M. (2024). Data-driven modeling techniques for prediction of settled water turbidity in drinking water treatment. *Frontiers in Environmental Engineering*, 3, 1401180. <https://doi.org/10.3389/fenve.2024.1401180>

- Mishra, R. K. (2023). Fresh Water availability and Its Global challenge. *British Journal of Multidisciplinary and Advanced Studies*, 4(3), 1–78. <https://doi.org/10.37745/bjmas.2022.0208>
- Narges, S., Ghorban, A., Hassan, K. & Mohamed, K. (2021) Predicción de la dosis óptima de coagulantes en plantas de tratamiento de agua mediante el desarrollo de modelos basados en el sistema de inferencia difusa de redes neuronales artificiales (ANFIS). *J Environ Health Sci Engineer* **19**, 1543–1553 <https://doi.org/10.1007/s40201-021-00710-0>
- Parmar, A., Katariya, R., & Patel, V. (2019). A Review on Random Forest: An Ensemble Classifier. In: Hemanth, J., Fernando, X., Lafata, P., Baig, Z. (eds) International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018. ICICI 2018. Lecture Notes on Data Engineering and Communications Technologies, vol 26. Springer, Cham. https://doi.org/10.1007/978-3-030-03146-6_86
- Salem, A. T., & AL-Musawi, N. O. (2021). Water Treatment With Conventional and Alternative Coagulants: A Review. *Journal of Engineering*, 27(9), 20-28. <https://doi.org/10.31026/j.eng.2021.09.02>
- Sawalkar, N.T., Jahav, S. W. y Pawa, A. A (2024) Prediction of Optimum Dosage of Coagulant in Water Treatment Plant: A Comparative Study between Artificial Neural Network and Random Forest. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2(05), 1408-1420. <https://doi.org/10.47392/IRJAEH.2024.0194>
- Sheng, D. P. W., Bilad, M. R., & Shamsuddin, N. (2023). Assessment and optimization of coagulation process in water treatment plant: A review. *ASEAN Journal of Science and Engineering*, 3(1), 79-100. <https://ejournal.kjpupi.id/index.php/ajse/article/view/319>
- Sun, D., Xu, J., Wen, H., & Wang, Y. (2020). An optimized random forest model and its generalization ability in landslide susceptibility mapping: application in two areas of Three Gorges Reservoir, China. *J. Earth Sci.*, 31, 1068-1086. <https://doi.org/10.1007/s12583-020-1072-9>
- Tahraoui, H., Belhadj, A. E., Moula, N., Bouranene, S., & Amrane, A. (2021). Optimisation and prediction of the coagulant dose for the elimination of organic micropollutants based on turbidity. *Kemija u industriji*, 70(11-12), 675-691. <https://doi.org/10.15255/KUI.2021.001>
- Tyralis H, Papacharalampous G, Langousis A. Una breve revisión de los bosques aleatorios para científicos y profesionales del agua y su historia reciente en recursos hídricos. *Water*. 2019; 11(5):910. <https://doi.org/10.3390/w11050910>
- Wadkar, D. V., Karale, R. S., & Wagh, M. P. (2021). Application of cascade feed forward neural network to predict coagulant dose. *Journal of Applied Water Engineering and Research*, 10(2), 87–100. <https://doi.org/10.1080/23249676.2021.1927210>
- Wang, D., Chen, L., Li, T., Chang, X., Ma, K., You, W., & Tan, C (2023) Successful prediction for coagulant dosage and effluent turbidity of a coagulation process in a drinking water treatment plant based on the Elman neural network and random forest models *Environ. Sci.: Water Res. Technol.*, 2023,9, 2263-2274 <https://doi.org/10.1039/D3EW00181D>
- Wu, J. Challenges for Safe and Healthy Drinking Water in China. *Curr Envir Health Rpt* 7, 292–302 (2020). <https://doi.org/10.1007/s40572-020-00274-5>
- Wu, J., Cao, M., Tong, D. et al. (2021) A critical review of point-of-use drinking water treatment in the United States. *npj Clean Water* **4**, 40 <https://doi.org/10.1038/s41545-021-00128-z>
- Yu, R. F., Kang, S. F., Liaw, S. L., & Chen, M. C. (2000). Application of artificial neural network to control the coagulant dosing in water treatment plant. *Water science and technology*, 42(3-4), 403-408. <https://doi.org/10.2166/wst.2000.0410>
- Zhang, K., Achari, G., Li, H., Zargar, A., & Sadiq, R. (2013). Machine learning approaches to predict coagulant dosage in water treatment plants. *International Journal of System Assurance Engineering and Management*, 4(2), 205–214. <https://doi.org/10.1007/s13198-013-0166-5>