



REVISTA TÉCNICA

DE LA FACULTAD DE INGENIERÍA

Una Revista Internacional Arbitrada
que está indizada en las publicaciones
de referencia y comentarios:

- SCIELO
- LATINDEX
- DOAJ
- MIAR
- REVENCYT
- AEROSPACE DATABASE
- CIVIL ENGINEERING ABTRACTS
- METADEX
- COMMUNICATION ABSTRACTS
- ZENTRALBLATT MATH, ZBMATH
- ACTUALIDAD IBEROAMERICANA
- BIBLAT
- PERIODICA

UNIVERSIDAD DEL ZULIA



REVISTA TÉCNICA
DE LA FACULTAD DE INGENIERÍA

Edificio Patrimonial La Ciega, primera sede de LUZ

“Post nubila phoebus”
“Después de las nubes, el sol”

LUZ en sus 130 años
de fundación
1891-2021

Modelamiento de Relaciones entre Parámetros Físicoquímicos y Microbiológicos en Aguas de la Bahía Interior del Lago Titicaca-Puno (Perú) mediante Árboles de Predicción

Edgar Eloy Carpio Vargas 

Universidad Nacional del Altiplano. Puno 21001. Peru
Autor de correspondencia: edgarcarpiovargas@gmail.com
<https://doi.org/10.22209/rt.v44n3a02>

Recepción: 14 de agosto de 2020 | Aceptación: 04 de junio de 2021 | Publicación: 01 de agosto 2021

Resumen

La cuantificación de parámetros físicoquímicos y microbiológicos es fundamental para definir la calidad del agua, pero el entendimiento de las relaciones y el comportamiento de estos parámetros, contribuyen a definir el estado trófico y calidad ambiental de los ecosistemas acuáticos. En este contexto, el objetivo del estudio fue determinar las relaciones entre los parámetros físicoquímicos y microbiológicos del agua de la bahía interior del lago Titicaca en la ciudad de Puno-Perú, aplicando la técnica de árboles de predicción de aprendizaje automático. La muestra estuvo constituida por 17 parámetros de calidad, que fueron medidos mensualmente durante enero de 2013 y abril de 2019, obteniéndose un total de 76 datos por cada parámetro. Para la elaboración de los árboles de predicción y su posterior interpretación, se utilizó el paquete estadístico R. Se evidenció que los parámetros físicoquímicos preponderantes del lago Titicaca, son temperatura, oxígeno disuelto, potencial de hidrógeno, conductividad eléctrica, sólidos suspendidos totales, clorofila-*a*, demanda bioquímica de oxígeno, demanda química de oxígeno y transparencia, a partir de los cuales se aplicó la técnica de árboles de predicción y se comprobó la eficiencia de estos modelos estadístico en la predicción de dichos parámetros para realizar estimaciones, teniendo en cuenta las características del lugar donde se utilice.

Palabras clave: aprendizaje automático; árbol de predicción; calidad de agua; modelamiento.

Modeling of Relationships between Physicochemical and Microbiological Parameters in Waters of the Interior Bay of Lake Titicaca-Puno (Peru) using Prediction Trees

Abstract

The quantification of physicochemical and microbiological parameters is fundamental to define water quality, but understanding the relationships and behaviour of these parameters contributes to define the trophic state and environmental quality of aquatic ecosystems. In this context, the objective of the study was to determine the relationships between the physicochemical and microbiological parameters of the water of the inner bay of Lake Titicaca in the city of Puno-Peru, applying the technique of machine learning prediction trees. The sample consisted of 17 quality parameters, which were measured monthly during January 2013 and April 2019, obtaining a total of 76 data for each parameter. For the elaboration of the prediction trees and their subsequent interpretation, the statistical package R was used. It was found that the main physicochemical parameters of Lake Titicaca are temperature, dissolved oxygen, hydrogen potential, electrical conductivity, total suspended solids, chlorophyll-*a*, biochemical oxygen demand, chemical oxygen demand and transparency, from which the technique of prediction trees was applied and the efficiency of these statistical models in the prediction of these parameters was checked to make estimates, taking into account the characteristics of the place where it is used.

Keywords: machine learning; modeling; prediction tree; water quality.

Introducción

El lago Titicaca se encuentra a una altitud de 3.810 msnm y se extiende en un área total de 8.400 km² aproximadamente (Chura y Mollocondo, 2009), además funciona como enlace entre Perú y Bolivia. Molina *et al.* (2017) mencionan que, al ser un recurso hídrico de alto potencial para los habitantes de sus islas y alrededores, proporciona servicios ecosistémicos como fuente acuifera, agua potable, riego, recreación, entre otros. No obstante, este lago no está exento de problemas de contaminación, según el ministerio del ambiente de Perú (MINAM, 2013), este lago se ve afectado por la contaminación proveniente de los relaves mineros, debido a que es colector de varios ríos; esta problemática podría provocar la alteración y una posible desaparición de la flora y fauna que lo habitan.

En 1980 las universidades de Columbia Británica de Canadá (UBC) y la Nacional de Tierras Altas de Puno (UNA), a través de sus estudios precisaron que la bahía interior de Puno es la zona más contaminada del lago de Titicaca, en una extensión aproximada a los 15 km² situada frente a la ciudad de Puno (Northcote *et al.*, 1989; Northcote *et al.*, 1991; Beltrán-Farfán *et al.*, 2015). Asimismo, un estudio realizado por Jiménez-Monroy *et al.* (2016), se evaluaron los parámetros fisicoquímicos en aguas residuales de la bahía de Puno confirmó que los valores de pH, conductividad eléctrica, sólidos totales disueltos, demanda bioquímica (DBO₅) y química de oxígeno (DQO), superan los límites permitidos. Pese a ello, la problemática en esta zona aún es latente en esta zona por lo que resulta de interés cuantificar la calidad del agua a través de modelos de predicción. Teniendo como alternativa los árboles de decisión o predicción, los cuales se definen como un modelo jerárquico que puede ser aplicado en un problema de regresión o clasificación. Es un modelo no paramétrico, debido a que el mismo no asume ninguna forma de distribución para las densidades de la variable. Su estructura se va generando durante el proceso y depende de la complejidad del problema inherente a los datos que se le proporcionen al modelo (Alpaydin, 2010). De acuerdo a lo anterior, un modelo para un árbol de predicción, se muestra como una estructura jerárquica donde se establecen las relaciones entre la variable dependiente y el conjunto de variables predictoras (Menacho-Chiok, 2017).

El funcionamiento básico de los árboles de predicción, está cimentado en el principio de los árboles de clasificación; donde la muestra de formación se ajusta mediante una división recursiva, es decir, se divide sucesivamente en subconjuntos cada vez más homogéneos, hasta que los nodos de la hoja contienen solo casos de una sola clase o se aplica algún otro criterio de parada razonable (Pompe y Feelders, 1997). El principal atractivo de estos modelos es su flexibilidad, ya que, su diseño no se basa en modelos preestablecidos, sino que lo van creando y ajustando su propio diseño, de tal manera que se ajustan sobre la marcha, según las particularidades del problema. Esta capacidad, los hace ideales para resolver problemas donde se cuenta con poca información, o están basados en variables cambiantes, para las cuales los métodos convencionales no están adaptados (Carbajo, 2016). Los modelos basados en metodologías de árboles de decisiones de aprendizaje automático, se han empleado en investigaciones para el modelamiento y la predicción de relaciones de propiedades y variables en diferentes procesos y entornos, como su uso para mejorar los resultados de experimentos fallidos (Raccuglia *et al.*, 2016) para la predicción de irradiancia (Castillo, 2015) y un estudio de relaciones respecto a los anuncios personalizados en tiendas en internet (Kim *et al.*, 2001).

En el caso de las relaciones entre las propiedades del agua, los trabajos realizados han centrado su interés en el uso de técnicas de correlación como las de Pearson y las predicciones mediante análisis de modelos lineales o modelos basados en geoestadística, así como en los modelos autorregresivos (De La Hoz y Gotilla, 2009; Carbajal y Sánchez, 2013; López *et al.*, 2015). Sin embargo, no se han encontrado estudios sobre el uso de modelos de árboles de predicción para relacionar los parámetros básicos del análisis del agua, por lo que en la presente investigación se planteó el uso de dichos modelos, para establecer las relaciones entre una serie de parámetros de calidad del agua del lago Titicaca en la ciudad de Puno en Perú, medidos entre 2013 y 2019.

Materiales y Métodos

Zona de estudio

La ciudad de Puno está ubicada en la región del mismo nombre, a orillas del lago Titicaca en Perú. Se ubica geográficamente en la sierra del sudeste, en la meseta del Collao entre los 13°01'00" y 17°17'30" de latitud sur y los 71°06'57" y 68°48'46" de longitud oeste del meridiano de Greenwich, con una altitud de 3.800 msnm (Gobierno Regional de Puno, 2020).

Población y muestra

Para el estudio, la base de datos se constituyó por una serie de parámetros fisicoquímicos y microbiológicos analizados en muestras de agua. Se analizaron 17 parámetros de calidad de agua diferentes, medidos mensualmente durante un periodo de 76 meses, entre enero de 2013 y abril de 2019, por personal del Laboratorio Continental del Instituto del Mar del Perú (IMARPE). Los parámetros utilizados para el estudio se describen en la Tabla 1.

Tabla 1. Descripción de los parámetros fisicoquímicos y microbiológicos del agua utilizados para el modelamiento.

Parámetro	Indicador	Símbolo	Unidad	Método	Referencia
Transparencia	Zona fótica	TA	m	Disco de Secchi	APHA <i>et al.</i> , 2000.
Temperatura	Temperatura	T	°C	Termometría	APHA <i>et al.</i> , 2000; Metcalf y Eddy, 1995.
pH	Grado de acidez o basicidad	pH	----	Electrométrico	APHA <i>et al.</i> , 2000; Metcalf y Eddy, 1995; OMS, 2006.
Conductividad eléctrica	Cantidad de iones inorgánicos	CE	μS/cm	Electrométrico	APHA <i>et al.</i> , 2000; Derisio, 1992; Metcalf y Eddy, 1995.
Demanda química de oxígeno	Contenido de materia orgánica	DQO	mg/L	Volumétrico	APHA <i>et al.</i> , 2000; Metcalf y Eddy, 1995.
Bacterias coliformes totales	Contaminación bacteriana	BCT	NMP /100 mL	Tubos múltiples	APHA <i>et al.</i> , 2000; CONAMA, 1986; EPA, 2021; OMS, 2006; LII 2021.
Bacterias coliformes termo tolerantes	Contaminación fecal	CTT	NMP /100 mL	Tubos múltiples	APHA <i>et al.</i> , 2000; CONAMA, 1986; EPA, 2021; OMS, 2006.
Clorofila- <i>a</i>	Productividad primaria	CF- <i>a</i>	mg/m ³	Espectrofotométrico	SIGAM, 2004 .
Demanda bioquímica de oxígeno	Contenido de materia orgánica biodegradable	DBO ₅	mg/L	Dilución	APHA <i>et al.</i> , 2000; Bueno <i>et al.</i> , 1997; Metcalf y Eddy, 1995.
Nitrato (NO ₃ ⁻)	Nivel trófico	NO ₃ ⁻	mg/L	Reducción de cadmio	Sawyer <i>et al.</i> , 2001.
Nitrito (NO ₂ ⁻)	Nivel trófico	NO ₂ ⁻	mg/L	Diazotización	Sawyer <i>et al.</i> , 2001.
Ortofosfato (PO ₄ ⁻³)	Nivel trófico	PO ₄ ³⁻	mg/L	Cloruro estañoso	CONAMA, 1986.
Fósforo (PT)	Nivel trófico	PT	mg/L	Cloruro estañoso	APHA <i>et al.</i> , 2000.
Nitrógeno amoniacal	Nivel trófico	N-NH ₃	mg/L	Nessler	APHA <i>et al.</i> , 2000.
Nitrógeno total (NT)	Nivel trófico	NT	mg/L	Nessler	APHA <i>et al.</i> , 2000; EPA, 2007.
Sólidos suspendidos totales	Turbidez del agua	SST	mg/L	Gravimetría	APHA <i>et al.</i> , 2000; Sawyer <i>et al.</i> , 2001.
Oxígeno disuelto	Relación producción/oxidación	OD	mg/L	Electrométrico	APHA <i>et al.</i> , 2000; Derisio, 1992; Metcalf y Eddy, 1995.

Análisis de datos

Para el modelamiento de las relaciones entre las variables, primero se realizó el análisis de normalidad multivariante (Mardia-tests y Royston) con significancia de 1 % ($\alpha=0,01$) y se tomó como premisa que $p>\alpha$, es decir, existe normalidad en todas las variables. Esto para realizar las pruebas de hipótesis de normalidad, además, se transformaron los datos y se corrigieron los valores atípicos, para así utilizar la estadística paramétrica y los modelos adecuados. Posterior a esto, se efectuó un análisis jerárquico de clúster. Finalmente, se aplicó la técnica no paramétrica de árboles de predicción de *machine learning*, que permite encontrar modelos de predicción (Management solutions, 2018; Sandoval, 2018; Rodrigo, 2020). Se usaron árboles Baggin y Random forest, aplicando el paquete estadístico R.

Resultados y Discusión

La evaluación *in situ* y de laboratorio registró algunos factores de contaminación debido a la fluctuación de ciertos parámetros físico-químicos de calidad del agua. Los valores obtenidos en los muestreos, se registran en la Tabla 2.

Tabla 2. Evaluación de parámetros fisicoquímicos y microbiológicos en el lago de Titicaca-Puno (Perú).

Parámetro	Indicador	Símbolo	Unidad	Media±DE
Transparencia	Zona fótica	TA	m	1,90±0,38
Temperatura	Temperatura	T	°C	16,00±2,01
pH	Grado de acidez o basicidad	pH	-	9,75±0,66
Conductividad eléctrica	Cantidad de iones inorgánicos	CE	µS/cm	1,78±3,23
Demanda química de oxígeno	Contenido de materia orgánica	DQO	mg/L	18,60±4,60
Bacterias coliformes totales	Contaminación bacteriana	BCT	NMP/100 mL	4.789,12±62,87
Bacterias coliformes termo tolerantes	Contaminación fecal	CTT	NMP/100 mL	3.080,90±139,01
Clorofila- <i>a</i>	Productividad primaria	CF- <i>a</i>	mg/m ³	90,00±6,48
Demanda bioquímica de oxígeno	Contenido de materia orgánica biodegradable	DBO ₅	mg/L	2,12±1,01
Nitrato	Nivel trófico	NO ₃ ⁻	mg/L	0,36±0,10
Nitrito	Nivel trófico	NO ₂ ⁻	mg/L	0,12±0,03
Ortofosfato	Nivel trófico	PO ₄ ³⁻	mg/L	1,38±0,56
Fósforo	Nivel trófico	PT	mg/L	2,43±0,26
Nitrógeno amoniacal	Nivel trófico	N-NH ₃	mg/L	1,12±0,19
Nitrógeno total	Nivel trófico	NT	mg/L	2,21±1,11
Sólidos suspendidos totales	Turbidez	SST	mg/L	953,00±0,25
Oxígeno disuelto	Relación producción/oxidación	OD	mg/L	7,80±1,23

DE: desviación estándar, $n=76$.

Prueba de normalidad multivariante

Para el análisis de la normalidad multivariante se ha utilizado el coeficiente de Mardia, en el que los valores inferiores a 0,70 se consideran indicativos de la normalidad multivariante de los datos (Tabla 3).

Tabla 3. Prueba de normalidad multivalente Mardia-tests y Royston.

Prueba	Estadístico	p	Conclusión
Mardia Skewness	25,918798	0,1710	Sí
Mardia Kurtosis	1,2973399	0,1953	Sí
MVN	NA	NA	Sí
Royston	7,151419	0,0602	Sí

MVN: prueba de Maren, p : nivel de significancia, NA: no aplica.

Las pruebas de Mardia y Royston indican que existe normalidad multivariante en los datos, ya que los valores probabilísticos son mayores al nivel de significancia $\alpha=0,01$. Según Bollen (2009), existe normalidad multivariada cuando el coeficiente de Mardia es inferior a $p(p+2)$, donde p es el número de variables observadas en el modelo, lo que al ser calculado da un valor de 323 y el coeficiente de Mardia obtuvo un valor, para el modelo corregido, de 74,33; y, en consecuencia, se puede afirmar que existe normalidad multivariada de los datos. Luego de observar las correlaciones entre las variables, se procedió al uso del análisis jerárquico de clúster. Considerando que, junto a los dendrogramas es una estrategia que funciona bajo una serie de principios que obligan a los investigadores a plantearse múltiples variables y considerar, en consecuencia, procedimientos de análisis multivariados para detectar el impacto que se tengan tanto en el proceso como en los actores (Marín, 2012). Además, como técnica multivariante permite agrupar los casos o las variables en función del parecido o la similitud entre ellos, ya que es un análisis, que partiendo de los elementos individualmente, va creando grupos hasta llegar a la formación de un único grupo o conglomerado constituido por todos los elementos de la muestra (Lin y Chen, 2006) (Figura 1).

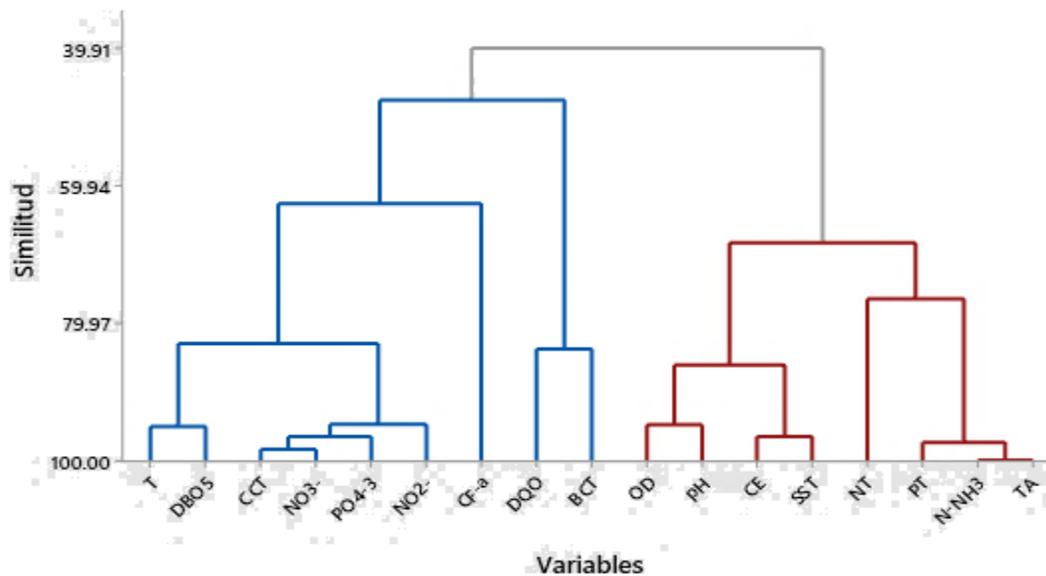


Figura 1. Dendrograma para el análisis jerárquico de clúster intergrupual k-medias, mostrando las relaciones entre los parámetros fisicoquímicos y microbiológicos de la bahía interior del lago Titicaca. La nomenclatura corresponde a la Tabla 1.

Haciendo un corte a una distancia re-escalada de 20, la combinación de clúster indica que las variables se asocian en dos conglomerados, un conglomerado 1 que contiene a OD, pH, CF-a, SST, DBO₅, PO₄³⁻, T y DQO, y un conglomerado 2 con las variables BCT, TA, N-NH₃, PT, NT, NO₃⁻, NO₂⁻, CE y CTT. Con base en lo anterior, se estableció un grupo de variables dependientes: T, OD, pH, CE, SST, CF-a, DBO₅, DQO y TA.

Modelamiento mediante árboles de predicción

Para todos los modelos, se utilizó un 80 % de los datos en el perfeccionamiento del modelo y 20 % para la validación. Según lo menciona Rodrigo (2017), los métodos estadísticos basados en árboles utilizan un conjunto de técnicas no paramétricas con las que se logra segmentar el espacio de los predictores en regiones simples, en las que se hace más sencillo manejar las interacciones entre las variables. En el caso de T, como variable dependiente, se muestra en la Figura 2.

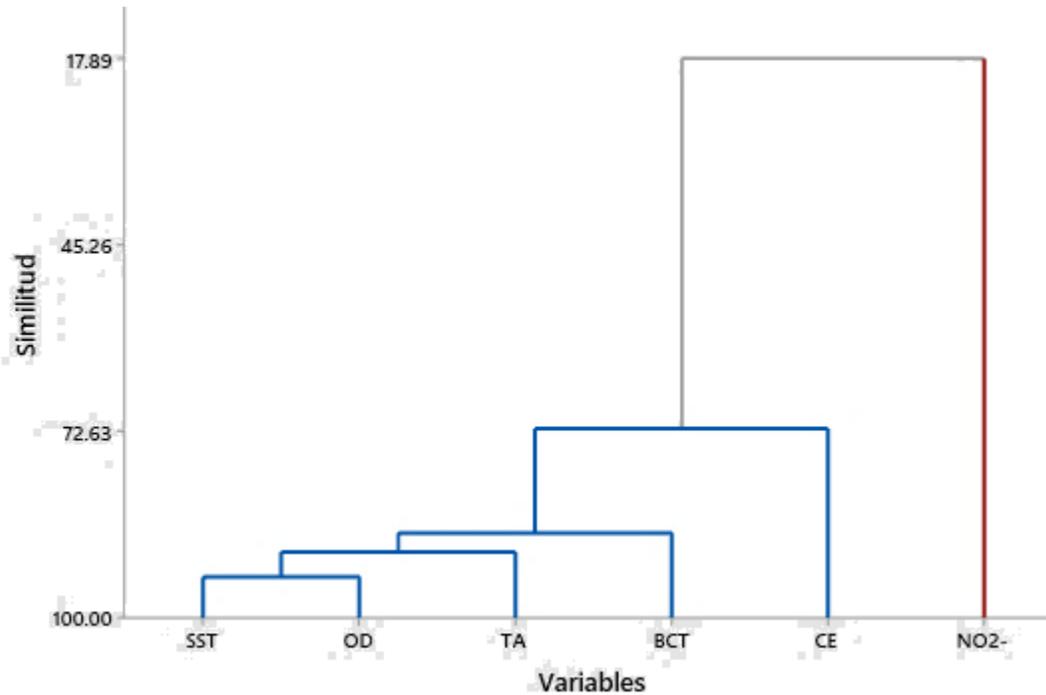


Figura 2. Árbol de predicción para establecer la relación entre la variable temperatura del agua y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

En la Figura 2, se observa que las variables más influyentes en la T del lago son: BCT, CE, OD, TA, SST y NO_2^- . En tanto, la interpretación del árbol de predicción indica, por ejemplo, que cuando se tienen valores de BCT > 665 NMP/100 mL y valores de OD $> 9,04$ mg/m^3 , la T alcanza $15,66$ $^{\circ}\text{C}$. Así, en este mismo escenario, si OD $< 9,04$ mg/m^3 , entonces la T muestra un valor de $12,71$ $^{\circ}\text{C}$.

La relación entre la T y los BCT no es concluyente para este estudio, no obstante, los resultados son similares a los reportados por Narváez *et al.* (2008) quienes, no evidencian relación directa entre la variable de T y BCT en un estudio realizado en la costa Caribe colombiana, cuyas características ecológicas, hidrológicas y geomorfológicas lo caracterizan como uno de los sistemas costeros más productivos del trópico; de la misma forma Gianoli *et al.* (2019), reportan que ciertos parámetros como T, tienen una significancia baja frente a las bacterias coliformes. Asimismo, se debe considerar que la T va a depender de otros factores, variando de acuerdo a la ubicación geográfica, las condiciones atmosféricas y climáticas, siendo en otros estudios un factor influyente en el crecimiento bacteriano (Okeola *et al.*, 2010; Nabi *et al.*, 2018; Haque *et al.*, 2019).

Para el OD, las variables más influyentes fueron N-NH_3 , CF-a, PO_4^{3-} , pH y DQO (Figura 3). Ciertos trabajos como el de Muñoz *et al.* (2015), reportan que el OD es un parámetro que se ve influenciado de manera importante por las demás variables del agua, como por ejemplo la salinidad, y que, además puede estar en función de la hora del día en que se tome la muestra para su determinación, por lo que es lógico su dependencia de los demás parámetros.

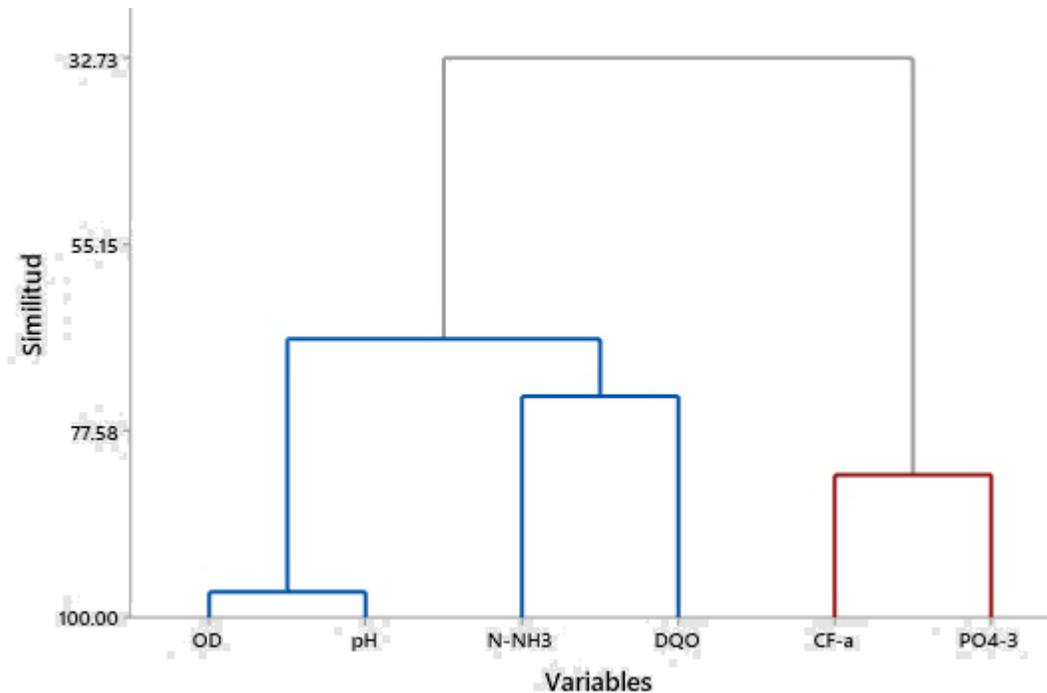


Figura 3. Árbol de predicción para establecer la relación entre la variable OD y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

Se observa entonces que para valores de $N-NH_3 > 1,30$ mg/L se predice un valor de OD de 6,05 mg/L. De igual manera, para valores $CF-a < 28,63$ mg/m³ y $N-NH_3 < 0,40$ mg/L se predice un valor de OD= 10,83 mg/L, siendo el mayor valor predicho de OD (11,40 mg/L), cuando se tiene $CF-a > 28,63$ mg/m³ y $PO_4^{3-} > 1,58$ mg/L. Se debe tener en cuenta que a menores valores de OD, se reduce la cantidad de iones y moléculas en el agua, siendo estos de importancia en la nutrición de microorganismos y otras formas de vida superior (Campaña *et al.*, 2017), su valor dependerá de la naturaleza del ecosistema, puesto que las variaciones del oxígeno disuelto son el resultado de las condiciones ambientales y biológicas (Brito *et al.*, 2016).

En cuanto a la variable pH, se obtuvo que los parámetros que más influyen fueron PT, NT, TA, BCT, T, PO_4^{3-} (Figura 4). Fadaeifard *et al.* (2012), afirman que el pH del agua está influenciado por las condiciones climáticas, y que establecer un rango adecuado del pH va a depender del uso que se le va dar al agua.

De acuerdo con la Figura 4, se puede interpretar que el mayor valor de pH= 10,19 se obtiene cuando el $PT < 1,10$ mg/L y el $NT < 0,85$ mg/L, y el menor valor predicho de pH= 8,05 se muestra para $PT > 1,10$ mg/L, $TA > 1,65$ m y $PO_4^{3-} > 1,18$ mg/L.

La CE mostró relación con los parámetros NT, PT, OD, CTT, y DBO_5 (Figura 5), debido a que está conformada por iones provenientes de sales disueltas y materia inorgánica componentes de los parámetros mencionados. Esto puede deberse a la presencia de otros aniones o cationes que también contribuyen a la conductividad eléctrica (Solís-Castro *et al.*, 2018).

Se observó que el valor máximo de $CE = 1,78$ $\mu S/cm$, se predice cuando $NO_3^- > 0,19$ mg/L, $OD < 9,02$ mg/L y $CTT > 14$ NMP/100 mL, y la mínima $CE = 1,52$ $\mu S/cm$ se predice cuando $NO_3^- < 0,19$ mg/L y $PT < 0,48$ mg/L. Los valores de CE concuerdan con los obtenidos por Beltrán-Farfán *et al.* (2015), cuyo estudio también fue realizado en el lago Titicaca.

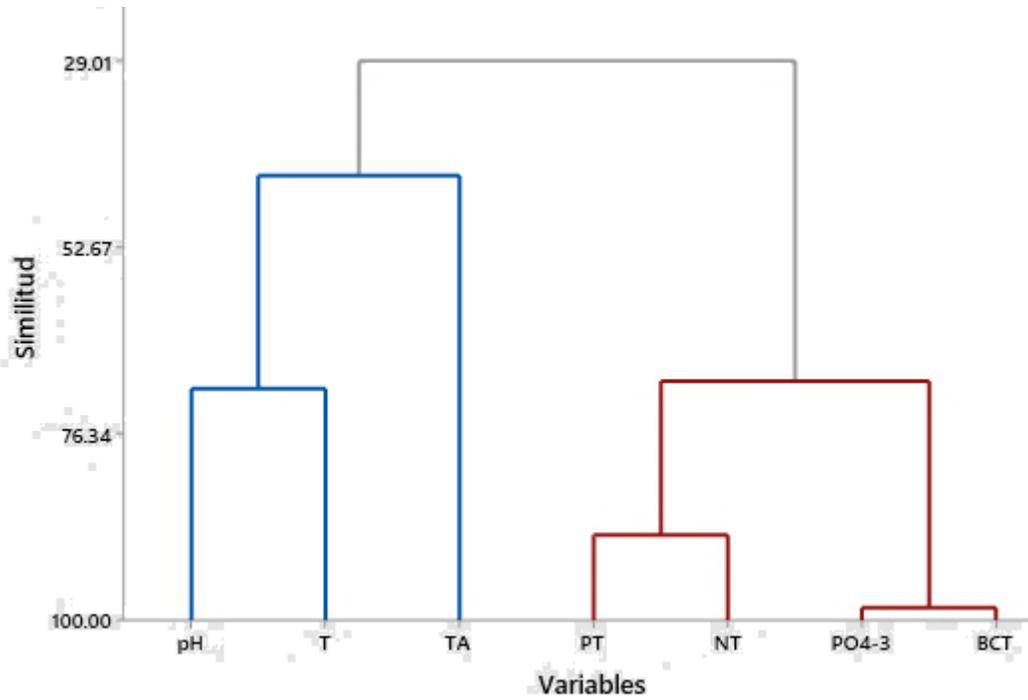


Figura 4. Árbol de predicción para establecer la relación entre la variable pH y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

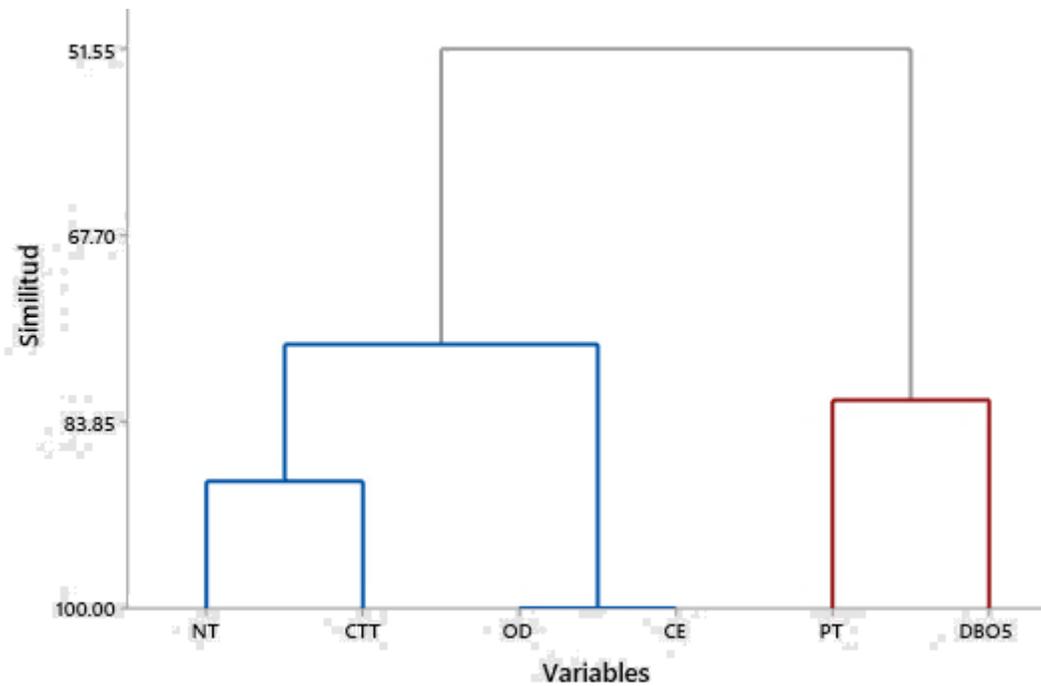


Figura 5. Árbol de predicción para establecer la relación entre la variable CE y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

El modelo para los SST (Figura 6) mostró que esta variable está relacionada con pH, OD, DBO₅, N-NH₃ y CF-a. Sánchez (2008) encontró que los SST estaban relacionados con la DQO, lo que no concuerda con lo obtenido en el presente estudio, sin embargo, en la investigación citada se utilizó un modelo de regresión lineal, lo cual difiere del enfoque aplicado en este estudio. Por su parte, Gutiérrez-Guzmán *et al.* (2014) reportaron altos niveles de SST y DBO₅ presentes en aguas residuales, ya que las altas concentraciones de SST en agua podrían bloquear la transferencia de oxígeno y que, el análisis de la DBO, es una prueba que se utiliza a nivel mundial para determinar las necesidades de oxígeno en la biodegradación de la materia orgánica presente en el agua (Isea *et al.*, 2015).

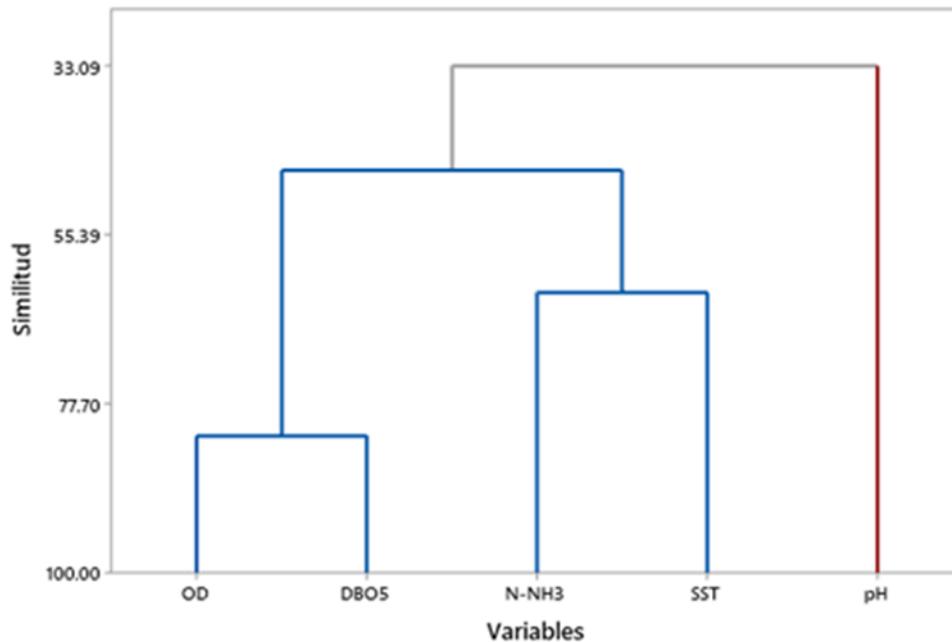


Figura 6. Árbol de predicción para establecer la relación entre la variable SST y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

Como se observa en la Figura 6, el árbol predice un valor mínimo de SST= 9,52 mg/L cuando el pH < 9,21 y un máximo de SST= 37,00 mg/L cuando pH > 9,21 y OD < 9,02 mg/L. A pesar de haber obtenido valores bajos de SST, se debe considerar que estos actúan como contaminantes a causa de la materia orgánica e inorgánica que la conforman (Mejía y Pérez, 2019).

La variable CF-a, mostró a través del árbol de predicción, relación con los parámetros pH, OD, T, CE, SST y PO₄³⁻ (Figura 7). La relación entre CF-a y PO₄³⁻ también fue observada al analizar agua del embalse del río Tercero en Argentina (Bonansea *et al.*, 2012), indicando que los valores de CF-a están supeditados no solo a la concentración de PO₄³⁻, sino también a factores como la T, la penetración de la luz y la locación.

Según el árbol de predicción, el valor mínimo de CF-a= 15,19 mg/m³ se predice cuando el pH < 9,32 y el OD < 8,58 mg/L, y el máximo de CF-a= 75,71 mg/m³ cuando pH > 9,32 y T < 15,90 °C. Según Bonansea *et al.* (2012) esta variable aumenta a lo largo del tiempo, sin embargo, los valores de CF-a dependen de las condiciones del entorno donde se realiza la medición.

En cuanto a la DBO₅, el árbol de predicción muestra que esta variable está relacionada con los parámetros NT, N-NH₃, OD, NO₃⁻, DQO y T (Figura 8). En cuanto a estas relaciones, una de las que se ha analizado es la reportada por Muñoz Nava *et al.* (2012), quienes, a través de un estudio realizado en la subcuenca del río Zahuapan en México, señalaron que la correlación entre la DBO₅ y la DQO mostró los valores más altos de r², obteniendo un valor general de 0,73 cuando se relaciona la carga al gasto de la DBO₅ con la población que habita hasta 20 km en la línea del cauce del río.

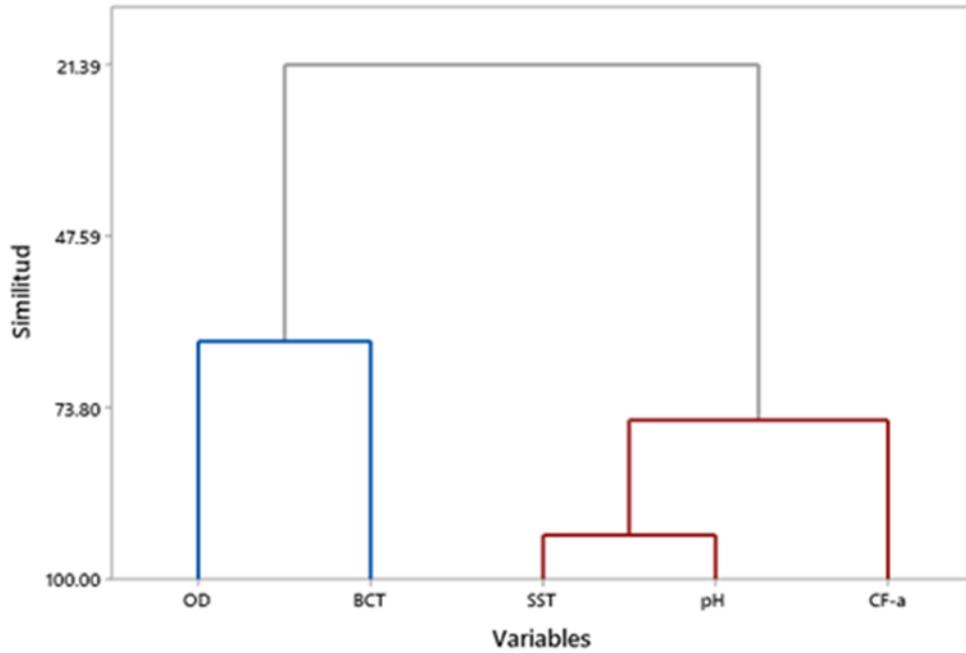


Figura 7. Árbol de predicción para establecer la relación entre la variable CF-a y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

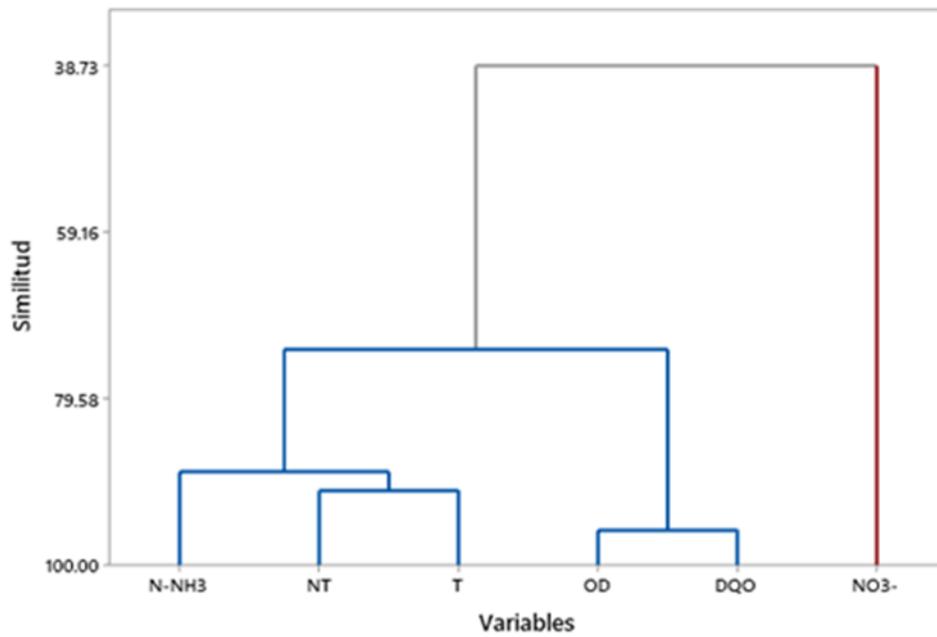


Figura 8. Árbol de predicción para establecer la relación entre la variable DBO₅ y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

A partir del análisis del árbol de predicciones se puede decir que el mínimo valor predicho para DBO₅= 4,71 mg/L se obtiene con NT > 2,19 mg/L, OD < 10,65 mg/L y DQO < 33,85 mg/L. Por otro lado, el máximo valor de DBO₅= 8,48 mg/L se predice cuando NT < 2,19 mg/L, N-NH₃ > 0,58 mg/L y NO₃⁻ > 0,15 mg/L.

En el análisis de árbol de predicción para la variable DQO, se obtuvo que la misma está relacionada con los parámetros CTT, T, CE, N-NH₃ y DBO₅ (Figura 9). Relaciones similares a las observadas en un estudio realizado en San Andrés (Ecuador), zona rural ubicada en el cantón Chone perteneciente a Manabí, una provincia rica en recursos naturales que posee numerosas áreas protegidas de bosques y reservas naturales; en un área aproximada de 200 km², situada en el km 11 de la vía Chone-Quito con una altitud de 45 msnm por Intriago y Cuenca-Nevárez (2014) reportaron una relación entre DQO y N-NH₃ consistente con los valores obtenidos en este estudio.

Se observa en la Figura 9 que el menor valor predicho de DQO= 21,02 mg/L se obtiene cuando CTT < 92,50 NMP/100 mL, T < 18,25 °C, N-NH₃ > 0,22 mg/L y 3,71 < DBO₅ < 8,87 mg/L. Por otro lado, el mayor valor de DQO= 40,60 mg/L se predice para CTT > 92,50 NMP/100 mL y CE < 1.754,02 µS/cm. Resultados contrarios a los reportados por Gil-Pavas *et al.*, (2018) a través de un estudio que mostraron una relación DBO₅/DQO de 0,2 en la etapa final.

En cuanto a la TA como variable, la misma se relaciona con los parámetros CF-a, pH, SST, NO₃⁻, NO₂⁻ y DQO (Figura 10). Al igual que lo observado, la transparencia también se ha relacionado de manera significativa con parámetros como SST y CF-a (Vásquez *et al.*, 2012).

El mayor valor de TA= 2,70 m se predice cuando CF-a < 11,60 mg/m³, y el menor valor de TA= 0,70 m cuando CF-a > 11,60 mg/m³; pH > 8,40; SST > 11,95 mg/L; NO₂⁻ > 0,10 mg/L y DQO < 31,90 mg/L. Según López *et al.* (2016), el árbol de predicción juega un papel muy importante para determinar la calidad del agua, ya que muchos sectores socioeconómicos dependen del uso de este preciado líquido.

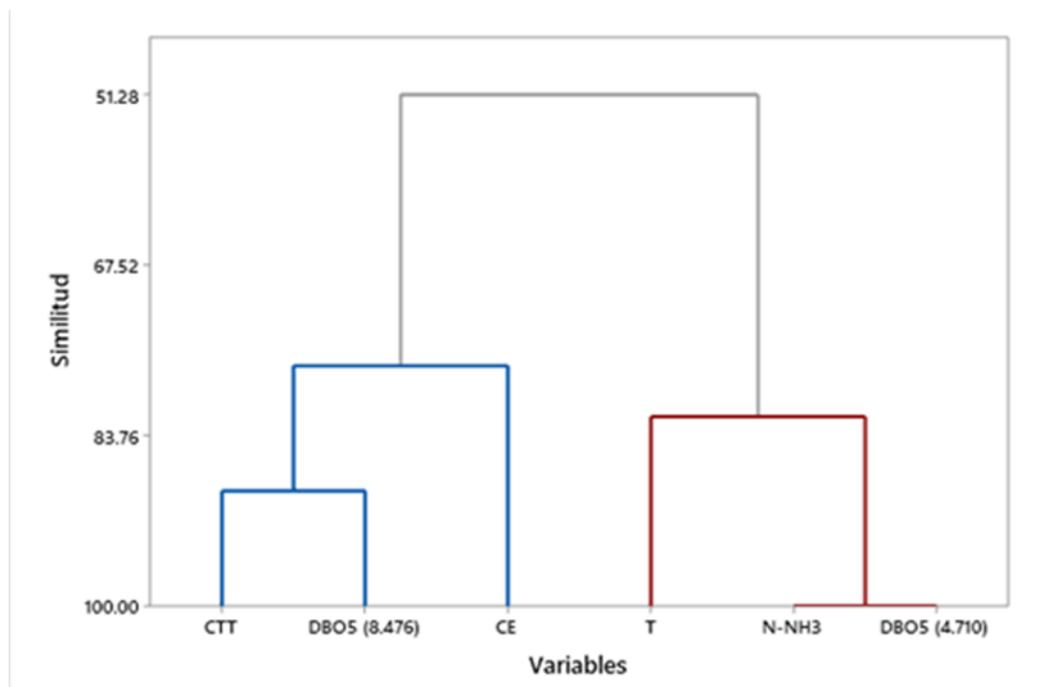


Figura 9. Árbol de predicción para establecer la relación entre la variable DQO y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

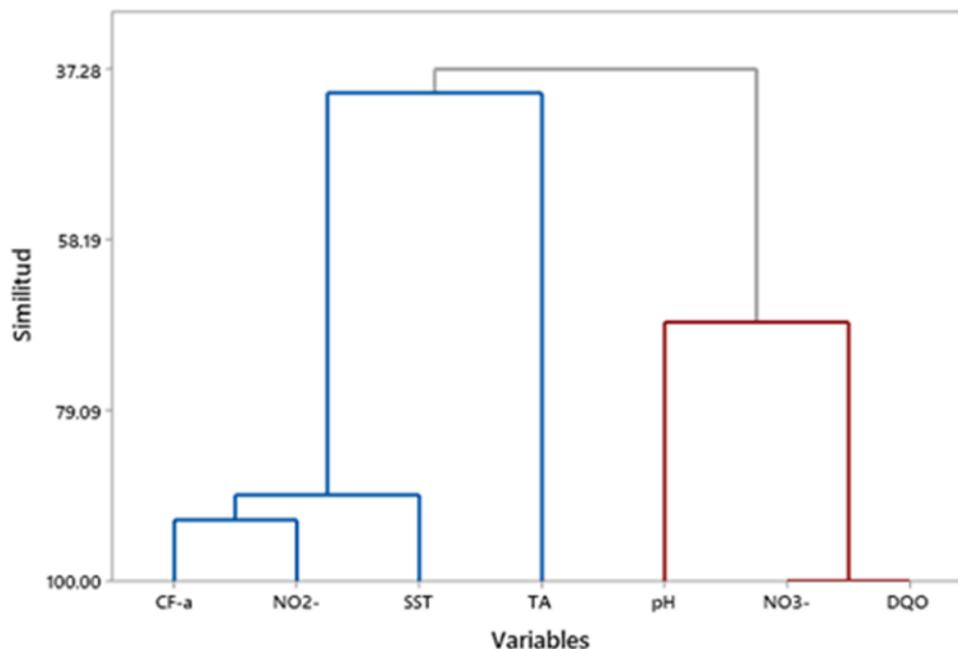


Figura 10. Árbol de predicción para establecer la relación entre la variable TA y los demás componentes del modelo. La nomenclatura corresponde a la Tabla 1.

Conclusiones

Con la aplicación del método del árbol de predicciones, se comprobó que las variables dependientes T, OD, pH, CE, SST, CF-a, DBO₅, DQO y TA, establecían sus propias relaciones con otros parámetros medidos en las aguas del lago Titicaca, llegando a la conclusión de que los modelos de árboles de predicciones permitían establecer vinculaciones entre los parámetros físicoquímicos y microbiológicos del agua, lo que los convierte en una herramienta útil para estimar los valores de una manera gráfica y fácil de interpretar, que bien puede considerarse un método general de evaluación. Sin embargo, el uso de esta metodología requiere tener en cuenta las diferencias entre los ecosistemas que se estudian, ya que cada uno tiene características y asociaciones diferentes.

Referencias Bibliográficas

- Alpaydin, E. (2010). *Introduction to machine learning*. Cambridge, Massachusetts: The MIT Press.
- APHA, AWWA, WPCF. (2000). *Métodos normalizados*. Madrid: Ediciones Díaz de Santos.
- Beltrán-Farfán, D., Palomino-Calli, R., Moreno-Terrazas, E., Peralta, C., Montesinos-Tubée, D. (2015). Calidad de agua de la bahía interior de Puno, lago Titicaca durante el verano del 2011. *Revista Peruana de Biología*, 22(3), 335-340.
- Bollen, K. (2009). *Structural equations with latent variables*. New York: Wiley.
- Bonanseña, M., Ledesma, C., Rodríguez, C., Sánchez, A. (2012). Concentración de clorofila-a y límite de zona fótica en el embalse río Tercero (Argentina) utilizando imágenes del satélite CBERS-2B. *Revista Ambiente & Agua*, 7(3), 61-71.
- Brito, D., Rivero, J., Guevara, M., Vásquez, F., Díaz, B., Gil, J. (2016). Análisis físicoquímico y microbiológico de la Laguna Grande, parroquia La Pica, Maturín - estado Monagas, Venezuela. *Saber*, 28(3), 502-510.
- Bueno, J., Sastre, H., Lavín, A. (1997). *Contaminación e ingeniería ambiental*. Oviedo: FICYT.

- Campaña, A., Gualoto, E., Chiluisa-Utreras, V. (2017). Evaluación físico-química y microbiológica de la calidad del agua de los ríos Machángara y Monjas de la red hídrica del distrito metropolitano de Quito. *Bionatura*, 2(2), 305-310.
- Carbajal, J., Sánchez, L. (2013). Diagnóstico y predicción del hábitat en la camaronicultura. *Computación y Sistemas*, 17(3), 435-455.
- Carbajo, I. (2016). *Experimentos computacionales en un estudio de simulación de modelos se series temporales para una mejor comprensión de las herramientas random forest y conditional trees*. Tesis de licenciatura. Madrid: Universidad Politécnica de Madrid.
- Castillo, N. (2015). *Técnicas de machine learning para el post-proceso de la predicción de la irradiancia*. Tesis de maestría. Granada: Universidad de Granada.
- Chura, R., Mollocondo, H. (2009). Desarrollo de la acuicultura en el lago Titicaca (Perú), *Revista AquATIC*, 31(1), 6-19.
- CONAMA. (1986). *Dispõe sobre a classificação das águas doces, salobras e salinas do território nacional*. Conselho Nacional do Meio Ambiente, Resolução N° 20 de 18/06/1986. Brasília.
- Intriago, F., Cuenca-Nevárez, G. (2014). Comportamiento de DBO₅, DQO, NH₄⁺ y NO₃⁻, mediante el diseño de un humedal artificial subsuperficial para depurar aguas residuales de origen doméstico. *La Técnica: Revista de las Agrociencias*, 13, 82-89.
- De La Hoz, C., Gotilla, C. (2009). Modelo para la predicción de parámetros de calidad del agua del lago de valencia utilizando imágenes de satélite. *XIII Congreso de la Asociación Española de Teledetección* (Teledetección, agua y desarrollo sostenible 2009). Ed. Salomón Montesinos Aranda y Lara Fernández Fornos.
- Derisio, J. (1992). *Introdução ao controle de poluição ambiental*. Sao Paulo: Oficina de textos.
- EPA. (2021). *Revised total coliform rule and total coliform rule* [en línea] disponible en: <https://www.epa.gov/dwreginfo/revised-total-coliform-rule-and-total-coliform-rule> [consulta 10 enero 2021].
- Fadaeifard, F., Raissy, M., Faghani, M., Majlesi, A., Nodeh, G. (2012). Evaluation of physicochemical parameters of waste water from rainbow trout fish farms and their impacts on water quality of Koohrang stream – Iran. *International Journal of Fisheries and Aquaculture*, 4(8). 170-177.
- Gianoli, A., Hung, A., Shiva, C. (2019). Relación entre coliformes totales y termotolerantes con factores físicoquímicos del agua en seis playas de la bahía de Sechura-Piura 2016-2017. *Salud y Tecnología Veterinaria*, 6(2), 62-71.
- Gil-Pavas, E., Arbeláez-Castaño, P., Medina-Arroyave, J., Gómez-Atehortua, C. (2018). Tratamiento de aguas residuales de la industria textil mediante coagulación química acoplada a procesos fenton intensificados con ultrasonido de baja frecuencia. *Revista Internacional de Contaminación Ambiental*, 34(1), 157-167.
- Gobierno Regional de Puno. (2020). *Información proyecto de presupuesto del sector público para el año 2021 - GORE Puno*. Gerencia Regional de Planeamiento, Presupuesto y Acondicionamiento Territorial.
- Gutiérrez-Guzmán, N., Valencia-Granada, E., Aragon-Calderon, R. (2014). Eficiencia de remoción en sistemas de tratamiento de aguas residuales del beneficio de café (*Coffea arabica*). *Colombia Forestal*, 17(2), 151-159
- Haque, M., Jewel, M., Sultana, M. (2018). Assessment of physicochemical and bacteriological parameters in surface water of Padma River, Bangladesh. *Applied Water Science*, 9(1), 1-8.

- Isea, D., Vargas, L., Durán, J., Delgado, J., Mendoza, R. (2021). Parámetros biocinéticos que rigen la ecuación de la DBO en aguas residuales de una industria procesadora de cangrejos. *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia*, 38(2), 112-121.
- Jimenez Monroy, L., Jahuira-Huarcaya, F., Ibañez-Quispe, V. (2016). Tratamiento de aguas eutrofizadas de la bahía interior de Puno, Perú, con el uso de dos Macrófitas. *Rev. Investig. Altoandina*, 18 (4), 403-410.
- Kim, J., Lee, B., Shaw, M., Chang, H., Nelson, M. (2001). Application of decision-tree induction techniques to personalized advertisements on internet storefronts. *International Journal of Electronic Commerce*, 5(3), 45-62.
- LII/Legal Information Institute. (2021). *40 CFR Part 136 - Guidelines establishing test procedures for the analysis of pollutants* [en línea] disponible en: <https://www.law.cornell.edu/cfr/text/40/part-136> [consulta 10 junio 2021].
- Lin, G., Chen, L. (2006). Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology*, 324(1-4),1-9.
- López, I., Figueroa, A., Corrales, J. (2016). Un mapeo sistemático sobre predicción de calidad del agua mediante técnicas de inteligencia computacional. *Revista Ingenierías Universidad de Medellín*, 15(28), 35-52.
- Management Solutions (2019). *Machine learning, una pieza clave en la transformación de los modelos de negocios* [en línea] disponible en: <http://bit.do/fF5JU> [consulta: 27 enero 2021].
- Marín, J. (2012). Los análisis clúster de tipo jerárquico y los dendogramas. Una visión para la triangulación metodológica en los estudios comparativos regionales en América Latina. Memorias. *Revista Digital de Historia y Arqueología desde el Caribe*, 39(1), 78-91.
- Menacho-Chiok, C. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26-33.
- Metcalf & Eddy, Inc. (1996). *Ingeniería de aguas residuales*. Ciudad de México: McGraw-Hill.
- MINAM. (2013). *Línea base ambiental de la cuenca del lago Titicaca*. Viceministerio de Gestión Ambiental.
- Molina, C., Lazzaro, X., Guédron, S., Achá, D. (2017). Contaminación de la bahía de Cohana, lago Titicaca (Bolivia): desafíos y oportunidades para promover su recuperación. *Ecología en Bolivia*, 52(2), 65-76.
- Muñoz, H., Orozco, S., Vera, A., Suárez, J., García, E., Neria, M., Jiménez, J. (2015). Relación entre oxígeno disuelto, precipitación pluvial y temperatura: río Zahuapan, Tlaxcala, México. *Tecnología y Ciencias del Agua*, 6(5), 59-74.
- Muñoz, H., Suárez, J., Vera, A., Orozco, S., Batlle, J., Ortiz, A., Mendiola, J. (2012). Demanda bioquímica de oxígeno y población en la subcuenca del río Zahuapan, Tlaxcala, México. *Revista Internacional de Contaminación Ambiental*, 28(1), 27-38.
- Nabi, B., Parveen, S., Hassan, T. (2018). Seasonal assessment of physicochemical parameters and evaluation of water quality of river Yamuna, India. *Advances in Environmental Technology*, 4(1), 41-49.
- Narváez, S., Gómez, M., Acosta, J. (2008). Coliformes termotolerantes en aguas de las poblaciones costeras y palafíticas de la ciénaga grande de Santa Marta, Colombia. *Acta Biológica Colombiana*, 13(3), 113-122.
- Northcote, T., Morales, P., Levy, D., Greaven, M. (1989). *Pollution in Lake Titicaca, Peru: training research and management*. Vancouver: University British Columbia.
- Northcote, T., Morales, P., Levy, D., Greaven, M. (1991). Contaminación en el lago Titicaca, Perú: Capacitación, investigación y manejo. Vancouver: University British Columbia.

Okeola, F., Kolawole, O., Ameen, O. (2010). Comparative study of physico-chemical parameters of water from a river and its surrounding wells for possible interactive effect. *Advances in Environmental Biology*, 4(3), 336-340.

Pompe, P., Feelders, A. (1997). Using machine learning, neural networks, and statistics to predict corporate Bankruptcy. *Computer-Aided Civil and Infrastructure Engineering*, 12(4), 267-276.

Raccuglia, P., Elbert, K., Adler, P., Falk, C., Wenny, M., Mollo, A., Zeller, M., Friedler, S., Schrier, J., Norquist, A. (2016). Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533(7601), 73-76.

Rodrigo, J. (2020). *Árboles de decisión, random forest, gradient boosting y C5.0* [en línea] disponible en https://www.cienciadedatos.net/documentos/33_arboles_de_prediccion_bagging_random_forest_boosting [consulta 9 junio 2021].

Rodrigo, J. (2017). «Clustering y heatmaps: aprendizaje no supervisado», RPubs. Recuperado de https://rpubs.com/Joaquin_AR/310338

Sánchez, J. (2008). *Estudio estadístico para la obtención de las relaciones necesarias entre parámetros analíticos de las aguas residuales para su caracterización según los modelos matemáticos de fangos activados*. Tesis de maestría. Valencia: Universidad Politécnica de Valencia

Sandoval, L. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. *Revista Tecnológica*, 11, 36-40.

Sawyer, C., McCarty, P., Parkin, G., Arteaga de García, L., Agudelo Quigua, D. (2001). *Química para ingeniería ambiental*. Ciudad de México: McGraw-Hill.

SIGAM. (2004). *Resumen ejecutivo agenda ambiental del municipio de Pasto*. Alcaldía Municipal de Pasto y Corporación Autónoma Regional de Nariño.

Solís-Castro, Y., Zúñiga-Zúñiga, L., Mora-Alvarado, D. (2018). La conductividad como parámetro predictivo de la dureza del agua en pozos y nacientes de Costa Rica. *Revista Tecnología en Marcha*, 31(1), 35-46

Vásquez, O., Pineda, S., Quiroga, E., Jara, B., Montiel, A. (2012). Relación entre clorofila-a y las variables oceanográficas en el área preglaciar del Seno Gallegos (Cordillera Darwin, Chile): bajo condiciones invernales. *Anales del Instituto de la Patagonia*, 40(1), 139-151.



UNIVERSIDAD
DEL ZULIA

REVISTA TECNICA

DE LA FACULTAD DE INGENIERIA
UNIVERSIDAD DEL ZULIA

Vol. 44. N°3, Septiembre - Diciembre, 2021 _____

*Esta revista fue editada en formato digital y publicada en Agosto 2021, por el **Fondo Editorial Serbiluz, Universidad del Zulia. Maracaibo-Venezuela***

www.luz.edu.ve
www.serbi.luz.edu.ve
www.produccioncientificaluz.org