

UNIVERSIDAD TECNOLÓGICA DE LOS ANDES

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA AMBIENTAL Y RECURSOS NATURALES



Tesis

Predicción de agentes patológicos mediante parámetros de calidad del agua con la aplicación de la inteligencia artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

Asesor:

Mg. Alvarado Ramos, Jessica

Autores:

Anampa Yalli, Rosy Sheila

Ancco Cervantes, Alex Junior

Para optar el Título Profesional: Ingeniero Ambiental

Abancay - Apurímac – Perú

2026

Acta de sustentación



UNIVERSIDAD TECNOLÓGICA DE LOS ANDES
FACULTAD DE INGENIERÍA AMBIENTAL
ESCUELA PROFESIONAL DE ING. AMBIENTAL Y RR.NN

Acta N°: 038

ACTA DE SUSTENTACIÓN DE TÍTULO PROFESIONAL

En la ciudad de Abancay, a los 14 días del mes de enero del 2026, siendo las 11:00 a.m. horas, se reunieron los integrantes del Jurado designado por Resolución Directoral N° 0258-2025-UTEA-FI-EPIARN de fecha 23 de diciembre del 2025 de la Escuela Profesional de Ing. Ambiental y Recursos Naturales, Facultad de Ingeniería:

| | |
|----------------|---------------------------------------|
| Presidente : | Mg. Salas Peña Vanesa |
| Dictaminante : | Mg. Sierra Puga Javier |
| Replicante : | Mg. Vargas Amiquero Milagros Carolina |

Para evaluar la sustentación, en la modalidad de:

Tesis Trabajo de suficiencia profesional

Titulada:

Predicción de agentes patológicos mediante parámetros de calidad del agua con la aplicación de la inteligencia artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

Desarrollado por el (los) Bachiller (es):

Br.: ANAMPA YALLI, Rosy Sheila
(Apellidos y Nombres)

Para optar el Título Profesional de:

Ingeniero Ambiental
(Denominación del Título)

Concluido el acto, el Jurado dictaminó que el (la) (los) mencionado(a) (s) bachiller (es) fue (ron) APROBADO (S):

Por: **Unanimidad**

(Unanimidad o Mayoría) (*)

Emitiéndose el calificativo final de:

| Bachiller (Apellidos y Nombres) | Calificación (**) |
|---------------------------------|-------------------|
| ANAMPA YALLI, Rosy Sheila | Aprobado |

Siendo las 12:30 horas concluyó la sesión, firmando los integrantes del Jurado.

Presidente: **MG. Salas Peña Vanesa**
(Dr. Mg.). (Apellidos y Nombres)

(Firma)

Dictaminante: **Mg. Sierra Puga Javier**
(Dr. Mg.). (Apellidos y Nombres)

(Firma)

Replicante: **Mg. Vargas Amiquero Milagros Carolina**
(Dr. Mg.). (Apellidos y Nombres)

(Firma)

(*): Mayoría: Dos integrantes del jurado aprueban o desaprueban; Unanimidad: Todos los integrantes del jurado aprueban o desaprueban, Art. 18 RGGAT.
(**): 0 a 10: Desaprobado, 11 a 15: Aprobado, 16 a 18: Aprobado Notable, 19 y 20: Aprobado con Distinción, Art. 18 RGGAT.



ACTA DE SUSTENTACIÓN DE TÍTULO PROFESIONAL

En la ciudad de Abancay, a los 14 días del mes de enero del 2026, siendo las 11:00 a.m. horas, se reunieron los integrantes del Jurado designado por Resolución Directoral N° 0258-2025-UTEA-FI-EPIARN de fecha 23 de diciembre del 2025 de la Escuela Profesional de Ing. Ambiental y Recursos Naturales, Facultad de Ingeniería:

| | |
|----------------|---------------------------------------|
| Presidente : | Mg. Salas Peña Vanesa |
| Dictaminante : | Mg. Sierra Puga Javier |
| Replicante : | Mg. Vargas Amiquero Milagros Carolina |

Para evaluar la sustentación, en la modalidad de:

- Tesis Trabajo de suficiencia profesional

Titulada:

Predicción de agentes patológicos mediante parámetros de calidad del agua con la aplicación de la inteligencia artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac - 2024.

Desarrollado por el (los) Bachiller (es):

Br.: ANCCO CERVANTES, Alex Junior
 (Apellidos y Nombres)

Para optar el Título Profesional de:

Ingeniero Ambiental
 (Denominación del Título)

Concluido el acto, el Jurado dictaminó que el (la) (los) mencionado(a) (s) bachiller (es) fue (ron) APROBADO (S):

Por: **Unanimidad**

(Unanimidad o Mayoría) (*)

Emitiéndose el calificativo final de:

| Bachiller (Apellidos y Nombres) | Calificación (**) |
|---------------------------------|-------------------|
| ANCCO CERVANTES, Alex Junior | Aprobado |

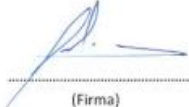
Siendo las 12:30 horas concluyó la sesión, firmando los integrantes del Jurado.

Presidente: **MG. Salas Peña Vanesa**
 (Dr. Mg.). (Apellidos y Nombres)



(Firma)

Dictaminante: **Mg. Sierra Puga Javier**
 (Dr. Mg.). (Apellidos y Nombres)



(Firma)

Replicante: **Mg. Vargas Amiquero Milagros Carolina**
 (Dr. Mg.). (Apellidos y Nombres)



(Firma)

(*): Mayoría: Dos integrantes del jurado aprueban o desaprueban; Unanimidad: Todos los integrantes del jurado aprueban o desaprueban, Art. 18 RGGAT.
 (**): 0 a 10: Desaprobado, 11 a 15: Aprobado, 16 a 18: Aprobado Notable, 19 y 20: Aprobado con Distinción, Art. 18 RGGAT.

Reporte de similitud






20% Similitud general

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

Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 10 palabras)

Fuentes principales

- 16%  Fuentes de Internet
- 7%  Publicaciones
- 15%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

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.



Metadatos

| Datos del autor | |
|--|---|
| Apellidos y Nombres | : Anampa Yalli, Rosy Sheila |
| Tipo de Documento de Identidad | : Documento Nacional de Identidad (DNI) |
| Número de Documento de Identidad | : 73338478 |
| URL ORCID | : https://orcid.org/0009-0003-4746-6019 |
| Apellidos y Nombres | : Ancco Cervantes, Alex Junior |
| Tipo de Documento de Identidad | : Documento Nacional de Identidad (DNI) |
| Número de Documento de Identidad | : 72899366 |
| URL ORCID | : https://orcid.org/0009-0005-6076-1463 |
| Datos del asesor | |
| Apellidos y Nombres | : Mg. Jessica Alvarado Ramos |
| Tipo de Documento de Identidad | : Documento Nacional de Identidad (DNI) |
| Número de Documento de Identidad | : 10372266 |
| URL ORCID | : https://orcid.org/0009-0003-2038-3117 |
| Datos de la investigación | |
| Facultad | : Ingeniería |
| Escuela Profesional | : Ingeniería Ambiental y Recursos Naturales |
| Línea de Investigación | : Calidad Ambiental |
| Rango de años en que se realizó la investigación | : 2024 - 2025 |
| Fuente de financiamiento | : Autofinanciado |
| Porcentaje de similitud | : 20% |
| URL de OCDE | : https://purl.org/pe-repo/ocde/ford#2.07.01 |

Dedicatoria

Primero, a Dios, por guiarme en cada paso y darme la fuerza para perseverar. A mi mamá Mercedes Candelaria Yalli Avendaño y Modesta Santusa Yalli Avendaño, por tu amor incondicional, apoyo constante y sacrificio incansable, que han sido mi mayor inspiración. Te amo con todo mi corazón, y esta tesis es mi humilde forma de agradecerte por todo lo que has hecho por mí.

Dedicatoria

Primeramente, a Dios por haberme permitido llegar hasta este punto con salud, sapiencia y las fuerzas necesarias para llevar a cabo este trabajo y a mis padres por el apoyo incondicional que me brindaron en toda la etapa de la elaboración de mi tesis.

Agradecimientos

- ✓ Agradecemos, en primer lugar, a nuestra casa de estudios Universidad Tecnológica de los Andes, docentes, administrativos; por su dedicación, comprensión y orientación, en las diferentes etapas de nuestra formación profesional.
- ✓ En segundo lugar, agradecemos al Mg. Yhon Huamán Fuentes quien nos guió durante toda la etapa de investigación; despertado nuestro interés por la investigación para aplicarlos en la vida y dar a conocer a la sociedad y mejorar la ciudad.
- ✓ Quiero agradecer también las críticas y el aporte intelectual realizaron a este trabajo mi familia, quienes sin su ayuda no se hubiera podido hacer esta tesis realidad.

Resumen

La presente investigación “Predicción de agentes patológicos mediante parámetros de calidad del agua con la aplicación de la inteligencia artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024” tuvo como objetivo determinar la predicción de los agentes patológicos mediante parámetros de la calidad del agua con la aplicación de la Inteligencia Artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024. La metodología utilizada fue tipo aplicada, enfoque cuantitativo, diseño no experimental. La población fueron los registros de monitoreo de calidad del agua existentes, tipo de muestra no probabilística – convencional con respecto a parámetros microbiológicos y parámetros físico-químicos; y muestreo por conveniencia. Los resultados obtenidos fueron según el coeficiente de determinación R^2 fue de 76,92%, lo que indica que, con un 95% de seguridad, la clasificación de los resultados tiene una aproximación mayor al 50%. Esto permitió aceptar la hipótesis alterna, sugiriendo que el uso de laboratorio e inteligencia artificial en los parámetros de calidad del agua muestra un grado de valoración $R^2 > 50\%$ en la determinación de agentes patológicos, lo que indica que existe una correlación positiva moderada y significativa donde las redes neuronales pueden predecir coliformes totales con seguridad del 95%. Se concluye que, el uso de laboratorio para el análisis de parámetros de calidad del agua e inteligencia artificial en los parámetros de calidad del agua, con un 95%, existe correlación positiva moderada esto significa que, la aplicación de inteligencia artificial en las redes neuronales para predecir agentes patológicos es viable.

Palabras Clave: Minería de datos; Aprendizaje automático; Predicción; Árbol de decisión; Análisis de regresión lineal.

Abstract

The present research “Prediction of Pathological Agents through Water Quality parameters with the application of Artificial Intelligence in the hydrographic basin of the Chumbao river, province of Andahuaylas, Apurímac region – 2024” aimed to determine the prediction of pathogenic agents using water quality parameters with the application of Artificial Intelligence in the Chumbao River Basin, Andahuaylas Province, Apurímac Region – 2024. The methodology used was applied, with a quantitative approach and a non-experimental design. The population consisted of existing water quality monitoring records, non-probabilistic sampling type – conventional with respect to microbiological and physicochemical parameters; and convenience sampling. The results obtained, according to the coefficient of determination R^2 , were 76.92%, indicating that, with 95% confidence, the classification of the results has an approximation greater than 50%. This allowed us to accept the alternative hypothesis, suggesting that the use of laboratory and artificial intelligence in water quality parameters shows a degree of accuracy $R^2 > 50\%$ in the determination of pathogenic agents, indicating a moderate and significant positive correlation where neural networks can predict total coliforms with 95% confidence. It is concluded that the use of laboratory analysis of water quality parameters and artificial intelligence in water quality parameters, with 95%, shows a moderate positive correlation. This means that the application of artificial intelligence in neural networks to predict pathological agents is viable.

Key words: Data Mining; Machine Learning; Prediction; Decision tree; Linear Regression Analysis.

Índice General

| | |
|---|------|
| Portada | i |
| Acta de sustentación | ii |
| Reporte de similitud | iv |
| Metadatos | v |
| Dedicatoria | vi |
| Agradecimientos | vii |
| Resumen (palabras clave) | viii |
| <i>Abstract (key words)</i> | ix |
| Índice General | x |
| Índice de Tablas | xiv |
| Índice de Figuras | xvi |
| Índice de Anexos | xix |
| I. Introducción..... | 20 |
| II. Planteamiento del Problema | 22 |
| 2.1 Descripción y formulación del problema..... | 22 |
| 2.1.1 Problema General | 24 |
| 2.1.2 Problemas Específicos..... | 24 |
| 2.2 Objetivos de la investigación..... | 25 |
| 2.2.1 Objetivo General | 25 |
| 2.2.2 Objetivo Específicos..... | 25 |
| 2.3 Justificación de la investigación | 25 |
| 2.4 Hipótesis..... | 26 |
| 2.4.1 Hipótesis General | 26 |

| | |
|--|----|
| 2.4.2 Hipótesis Específicos..... | 26 |
| 2.5 Variables | 27 |
| 2.5.1 Variable 1 | 27 |
| 2.5.2 Variable 2 | 27 |
| 2.6 Operacionalización de Variables | 28 |
| III. Marco Teórico..... | 30 |
| 3.1 Antecedentes de la investigación..... | 30 |
| 3.1.1 A nivel internacional | 30 |
| 3.1.2 A nivel nacional..... | 33 |
| 3.1.3 A nivel regional y local | 36 |
| 3.2 Bases teóricas | 38 |
| 3.2.1 El agua..... | 38 |
| 3.2.2 Calidad de agua | 38 |
| 3.2.3 La contaminación del agua | 39 |
| 3.2.4 Características físico-químico del agua. | 41 |
| 3.2.5 Características microbiológicas del agua..... | 52 |
| 3.2.6 Microorganismos bioindicadores..... | 58 |
| 3.2.7 Factores que intervienen en la transmisión de enfermedades..... | 59 |
| 3.2.8 Principales enfermedades por el consumo de aguas residuales..... | 60 |
| 3.2.9 Inteligencia Artificial..... | 66 |
| 3.2.10 Aprendizaje automático (Machine Learning) | 67 |
| 3.2.11 Análisis de datos (Data Analysis) | 69 |

| | |
|--|-----|
| 3.2.12 Redes neuronales (Neural Networks) | 70 |
| 3.2.13 Aprendizaje profundo (Deep Learning)..... | 76 |
| 3.2.14 Métricas de evaluación | 78 |
| 3.2.15 Marco legal..... | 81 |
| 3.3 Definición de Términos..... | 87 |
| IV. Metodología..... | 89 |
| 4.1 Tipo y nivel de investigación..... | 89 |
| 4.1.1 Tipo de investigación | 89 |
| 4.1.2 Nivel de investigación | 89 |
| 4.2 Diseño de investigación..... | 90 |
| 4.3 Ámbito temporal y espacial..... | 90 |
| 4.3.1 Población y muestra | 90 |
| 4.4 Instrumentos..... | 98 |
| 4.5 Procedimientos..... | 99 |
| 4.5.1 Diagrama de flujo del proyecto de investigación..... | 99 |
| 4.6 Análisis de datos..... | 116 |
| 4.6.1 Procedimiento Estadístico | 116 |
| V. Resultados y Discusión | 131 |
| 5.1 Resultados para el objetivo general | 131 |
| 5.2 Resultados para los objetivos específicos | 134 |
| 5.2.1 O.E.1: Resultados del objetivo específico 01..... | 134 |
| 5.2.2 O.E.2: Resultados del objetivo específico 02..... | 135 |

| | |
|---|-----|
| 5.2.3 O.E.3: Resultados del objetivo específico 03..... | 136 |
| 5.3 Discusión de Resultados..... | 137 |
| VI. Conclusiones..... | 141 |
| VII. Recomendaciones..... | 143 |
| VIII. Referencias..... | 145 |
| IX. Anexos..... | 164 |

Índice de Tablas

| | |
|--|-----|
| Tabla 1 | 28 |
| Matriz de la Operacionalización de Variables | 28 |
| Tabla 2 | 83 |
| Categorías en relación al uso para evaluar los parámetros según ECA Agua | 83 |
| Tabla 3 | 86 |
| Categoría 1. Poblacional y recreacional, Sub Categoría A1 – ECA | 86 |
| Tabla 4 | 91 |
| Puntos de muestreo | 91 |
| Tabla 5 | 100 |
| Datos químicos y biológicos | 100 |
| Tabla 6 | 105 |
| Métricas en la Regresión Lineal | 105 |
| Tabla 7 | 107 |
| Métricas de Evaluación en Decision Tree | 107 |
| Tabla 8 | 109 |
| Métricas de Evaluación en Random Forest | 109 |
| Tabla 9 | 109 |
| Primer Modelo de Red Neuronal Artificial | 109 |
| Tabla 10 | 112 |
| Segundo Modelo de ANN | 112 |
| Tabla 11 | 117 |
| Datos de entrada para los modelos de aprendizaje | 117 |
| Tabla 12 | 117 |
| Datos de entrada para los modelos de aprendizaje | 117 |
| Tabla 13 | 121 |
| Resultados de la aplicación del modelo de Red Neuronal | 121 |

| | |
|---|-----|
| Tabla 14 | 122 |
| Datos Estadísticos O.E.1 | 122 |
| Tabla 15 | 124 |
| Resultados de la aplicación del modelo de Regresión Lineal | 124 |
| Tabla 16 | 125 |
| Resultados de la aplicación del modelo de redes neuronales artificiales | 125 |
| Tabla 17 | 126 |
| Datos Estadísticos O.E.2 | 126 |
| Tabla 18. | 165 |
| Matriz de consistencia..... | 165 |
| Tabla 19. | 169 |
| Matriz de la Operacionalización de Variables | 169 |

Índice de Figuras

| | |
|--|-----|
| Figura 1 | 44 |
| Materia orgánica oxidada a bióxido de carbono y agua..... | 44 |
| Figura 2 | 45 |
| Seguimiento de una planta de tratamiento de aguas residuales | 45 |
| Figura 3 | 58 |
| Principales microorganismos bioindicadores | 58 |
| Figura 4 | 69 |
| Tipos de aprendizaje automático | 69 |
| Figura 5 | 72 |
| Modelo de red neuronal | 72 |
| Figura 6 | 74 |
| Estructura de redes neuronales convolucionales (CNN)..... | 74 |
| Figura 7 | 75 |
| Estructura de redes neuronales recurrentes RNN | 75 |
| Figura 8 | 78 |
| Estructura de aprendizaje profundo..... | 78 |
| Figura 9 | 91 |
| Población de la cuenca del río Chumbao..... | 91 |
| Figura 10 | 97 |
| Arquitectura Propuesta de una Red Neuronal..... | 97 |
| Figura 11 | 99 |
| Diagrama de flujo del proyecto de predicción de Coliformes Totales | 99 |
| Figura 12 | 101 |
| Puntos de muestra de nuestros datos. | 101 |
| Figura 13 | 102 |
| Datos de entrenamiento y testeo para los modelos. | 102 |

| | |
|---|-----|
| Figura 14 | 103 |
| Datos escalados para su entrenamiento | 103 |
| Figura 15 | 104 |
| Modelo de regresión lineal | 104 |
| Figura 16 | 106 |
| Modelo Decision Tree | 106 |
| Figura 17 | 108 |
| Modelo Random Forest | 108 |
| Figura 18 | 110 |
| Primer Modelo de ANN | 110 |
| Figura 19 | 111 |
| Primer Modelo de ANN - Validation Loss | 111 |
| Figura 20 | 113 |
| Segundo Modelo de ANN - Validation Loss | 113 |
| Figura 21 | 114 |
| Primer Modelo de ANN regresión lineal en los datos de Training y testeo | 114 |
| Figura 22 | 115 |
| Primer Modelo de ANN regresión lineal en los datos reales y predichos | 115 |
| Figura 23 | 116 |
| Segundo Modelo de ANN regresión lineal en los datos reales y predichos | 116 |
| Figura 24 | 119 |
| Predicción de coliformes totales mediante regresión lineal | 119 |
| Figura 25 | 119 |
| Predicción de coliformes totales mediante Random Forest | 119 |
| Figura 26 | 120 |
| Predicción de coliformes totales mediante Redes Neuronales Artificiales | 120 |
| Figura 27 | 123 |

| | |
|--|-----|
| Regresión lineal cálculo de R^2 | 123 |
| Figura 28..... | 128 |
| Regresión lineal del modelo de aprendizaje supervisado cálculo de R^2 | 128 |
| Figura 29..... | 129 |
| Regresión lineal del modelo de redes neuronales artificiales, cálculo de R^2 | 129 |
| Figura 30..... | 131 |
| Matriz de confusión de los parámetros de la calidad del agua..... | 131 |
| Figura 31..... | 132 |
| Resultado del Modelo de Regresión Lineal..... | 132 |
| Figura 32..... | 133 |
| Resultado del Modelo de Random Forest..... | 133 |
| Figura 33..... | 135 |
| Mean Absolute Error (MAE) - Red Neuronal Artificial..... | 135 |
| Figura 34..... | 136 |
| Resultado del Modelo de Redes Neuronales Artificiales..... | 136 |

Índice de Anexos

| | |
|--|-----|
| Anexo 1. Matriz de consistencia | 164 |
| Anexo 2. Operacionalización de Variables | 168 |
| Anexo 3. Instrumento de recolección de información | 171 |
| Anexo 4. Panel fotográfico (Visita a diferentes puntos del río Chumbao) | 174 |
| Anexo 5. Recolección de Datos de Diferentes Fuentes | 185 |

I. Introducción

El título del estudio es predicción de agentes patológicos mediante parámetros de calidad del agua con la aplicación de la inteligencia artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024. Con una línea de investigación de calidad ambiental.

El agua es un recurso vital, esencial para la vida en todo el planeta. Aunque se renueva naturalmente, es finito y crucial para el desarrollo de todos los seres vivos. La contaminación del agua tiene un impacto directo y perjudicial en la salud de los seres humanos, la flora, la fauna y los ecosistemas en general. (Baquerizo y Acuña, 2019)

La Inteligencia Artificial ha permitido la creación de sistemas avanzados capaces de simular actividades complejas, como la predicción de agentes patológicos a partir del análisis de parámetros de calidad del agua. Estos sistemas utilizan algoritmos de machine learning y análisis de datos para identificar patrones, detectar anomalías y prever riesgos de contaminación, lo cual es fundamental para la salud pública y la gestión ambiental.

Según la Organización Mundial de la Salud (OMS) los agentes patógenos transmitidos por el agua constituyen una dificultad de salud ambiental, frente a esta problemática existe una demanda urgente de control mediante la implementación de medidas de protección ambiental, a fin de evitar el incremento de la prevalencia de las enfermedades relacionadas con la calidad del agua de acuerdo a su procedencia y almacenamiento, considerando el coste beneficio y su efectividad. Por ello es necesario el conocimiento de las características fisicoquímicas y microbiológicas del agua de consumo para cumplir con las exigencias de la entidad reguladora (OMS, 1998).

Capítulo I, la introducción desarrolló de manera general las variables de la investigación y desglosó de manera corta cada capítulo del presente informe.

Capítulo II, el planteamiento del problema se desarrolló acorde a la ciudad de Andahuaylas y el río Chumbao, puesto que comparten el mismo espacio en la cuenca. La ciudad no ha sido capaz de establecer una dinámica positiva de coexistencia con el río lo cual podría provocar problemas de salud en la población. (Aguirre y Diaz, 2018).

Esta investigación se desarrolló en la región Apurímac – Perú, específicamente en lo largo de la cuenca hidrográfica del río Chumbao, la cual comprende diferentes comunidades aledañas a la cuenca Chumbao de Andahuaylas.

Capítulo III, la fuente de información recaudada para el marco teórico sirvió para comparar resultados de este estudio mediante antecedentes internacionales, nacionales y locales logrando coadyuvar a investigaciones, más base teórica, marco normativo y términos básicos.

Capítulo IV, la metodología del estudio es experimental, cuya técnica es la observación directa y base de datos, aplicando el método de regresión lineal utilizando el programa Minitab., en el ámbito de la cuenca Chumbao.

Capítulo V, en cuanto a los resultados y discusión, se aplicó el método de regresión lineal a través del programa Minitab, en el ámbito la cuenca Chumbao donde se observó claramente la valoración $R^2 > 50\%$ en la determinación de agentes patológicos, lo que implica que las redes neuronales pueden predecir coliformes totales con seguridad del 95%. Los cuales se discutieron con algunos de nuestros antecedentes identificados que llevaban relación de resultados.

II. Planteamiento del Problema

2.1 Descripción y formulación del problema

En el mundo entero, la contaminación de las aguas superficiales lóxicas, se debe principalmente al incremento poblacional de ciudades cercanas a estas fuentes y el incremento de generación per cápita de los residuos sólidos, ocasionando la alteración de su composición natural, la afectación de la biodiversidad y a los seres humanos que se abastecen de estos cuerpos de agua (alimentación, recreación, pesca, actividades agrícolas, etc.). Según (Conagua, 2018), el 59.1% de los ríos, arroyos, lagos, lagunas, presas y zonas costeras monitoreadas están contaminados.

En el Perú, más de diez millones de peruanos, el 31,15% de la población ingiere agua contaminada. Por lo tanto, la gestión adecuada y conservación de los recursos hídricos es uno de los desafíos más significativos que enfrenta nuestra nación. Debido a que el agua es un recurso limitado y es fundamental para el desarrollo de la vida humana. (Arrojo, 2022)

En la región Apurímac, la disponibilidad de agua es una gran preocupación ya que, se encuentra contaminada por desechos fecales y químicos. En las áreas urbanas, el 25.2% de la población accede al servicio sin contaminación por restos fecales y químicos, mientras que, en las áreas rurales, el 1.9% de la población usa agua contaminada. (Sunass, 2022)

En la cuenca del río Chumbao, los pobladores aledaños enfrentan un grave problema de calidad del agua debido a la presencia de agentes patológicos, es decir, numerosos análisis microbiológicos han confirmado que los niveles de estos contaminantes superan los Estándares de Calidad Ambiental (ECA) establecidos por la normativa peruana, como el Decreto Supremo N° 004-2017-MINAM. Análisis que muchas veces se ven limitados por

los altos precios de los laboratorios ya que en la misma localidad no se encuentran laboratorios acreditados para su análisis.

Las actividades diarias de sus habitantes de las zonas urbanas de Andahuaylas, San Jerónimo y Talavera de la Reyna aportan a la contaminación por sus mismos desperdicios desechados a las orillas del río (compuesto por todo tipo de residuos sólidos). Escorrentías agrícolas y ganaderas, que transportan hacia el río desechos animales, fertilizantes y pesticidas representando un riesgo directo para la salud pública. De esta manera, de los 1329 casos, el 24.7% presenta síntomas y signos relacionados con el sistema digestivo y el abdomen, y del 10% de los 537 casos, una infección aguda de las vías respiratorias superiores.

En consecuencia, la contaminación microbiológica, manifestada en la presencia de coliformes fecales (como la *E. coli*) que superan los Estándares de Calidad Ambiental (ECA) para agua, harán de este recurso inadecuado para consumo humano, riego y otras actividades vitales, lo que deriva en potenciales enfermedades gastrointestinales y otros problemas de salud en las comunidades locales.

Es por ello que, el presente estudio propone la implementación de la Inteligencia Artificial específicamente mediante la regresión lineal en el software Minitab, para predecir agentes patológicos en el agua usando parámetros físico-químicos. El objetivo es ofrecer un método más eficiente y económico que los análisis tradicionales para estimar la contaminación, facilitando la gestión de la calidad del agua y la toma de decisiones oportuna en la cuenca del río Chumbao.

Bajo estos argumentos se ve la necesidad de investigar sobre la predicción de agentes patológicos mediante parámetros de calidad del agua con la aplicación de la inteligencia artificial en la cuenca hidrográfica del río Chumbao.

2.1.1 Problema General

- ✓ ¿Cómo predecir los agentes patológicos mediante parámetros de la calidad del agua con la aplicación de la Inteligencia Artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024?

2.1.2 Problemas Específicos

- ✓ ¿Cuál es el nivel de correlación entre los valores reales de los parámetros microbiológicos y parasitológicos de la calidad del agua y los valores predichos por la aplicación de la inteligencia artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024?
- ✓ ¿Cuál de los modelos de inteligencia artificial, entre Neuronal Network o Machine Learning Regression presenta mayor precisión en la predicción de los parámetros microbiológicos y parasitológicos en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024?
- ✓ ¿En qué medida los parámetros físico químicos del agua pueden ser utilizados como información de datos de entrada en los modelos de inteligencia artificial para determinar la predicción de los parámetros microbiológicos y parasitológicos en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024?

2.2 Objetivos de la investigación

2.2.1 Objetivo General

- ✓ Determinar la predicción de los agentes patológicos mediante parámetros de la calidad del agua con la aplicación de la Inteligencia Artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

2.2.2 Objetivo Específicos

- ✓ Evaluar el nivel de correlación entre los valores reales de los parámetros microbiológicos y parasitológicos de la calidad del agua y los valores predichos por la aplicación de la inteligencia artificial en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.
- ✓ Evaluar qué modelo de inteligencia artificial, entre Neuronal Network o Machine Learning Regression presenta mayor precisión en la predicción de los parámetros microbiológicos y parasitológicos en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.
- ✓ Determinar en qué medida los parámetros físico químicos del agua pueden ser utilizados como información de datos de entrada en los modelos de inteligencia artificial para determinar la predicción de los parámetros microbiológicos y parasitológicos en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

2.3 Justificación de la investigación

El trabajo de investigación actual nos permitirá lograr una mejor comprensión de la calidad en el que se encuentra el río Chumbao, predecir y determinar los agentes patológicos que se encuentran en él, haciendo uso de la inteligencia artificial.

La implementación de IA nos permitirá predecir agentes patológicos; y para lo cual proponemos modelos híbridos de aprendizaje como Neuronal Network y Machine Learning Regression que daremos uso en aplicación de la Inteligencia Artificial para determinar agentes patológicos.

El presente estudio será beneficioso para toda la población al proporcionar información y servir como base para investigaciones futuras sobre las diversas cuencas de la región Apurímac y otras áreas.

2.4 Hipótesis

2.4.1 Hipótesis General

- ✓ Aplicando la Inteligencia Artificial si se predicen agentes patológicos mediante parámetros de la calidad del agua en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

2.4.2 Hipótesis Específicos

- ✓ La correlación entre los valores reales de los parámetros microbiológicos y parasitológicos de la calidad del agua y los valores predichos por la aplicación de la inteligencia artificial presenta un grado de correlación $R^2 > 50\%$ en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.
- ✓ El modelo de inteligencia artificial Neuronal Network presenta mayor precisión en la predicción de los parámetros microbiológicos y parasitológicos con respecto al Machine Learning Regression en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

- ✓ Los parámetros físicos químicos del agua se aplican como datos de entrada en los modelos de inteligencia artificial para determinar la predicción de los parámetros microbiológicos y parasitológicos en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

2.5 Variables

2.5.1 Variable 1

Predicción de los Agentes Patológicos mediante parámetros de la calidad del agua. El análisis predictivo se fundamenta en la recopilación e interpretación de datos previos para estimar tendencias y comportamientos futuros, cuya precisión depende de la dinámica y variabilidad del fenómeno estudiado. En estudios ambientales y sanitarios, este proceso resulta clave para evaluar la presencia de agentes patológicos microorganismos con capacidad de producir enfermedad y para determinar la calidad del agua, medida a través de parámetros físicos, químicos y biológicos que establecen su idoneidad para la vida y el consumo humano. (Makridakis, 1997) (Murray, 2020) (Organización Mundial de la Salud, 2017)

2.5.2 Variable 2

Aplicación de la Inteligencia Artificial. La Inteligencia Artificial (IA) es la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano. Sin embargo, a diferencia de las personas, los dispositivos basados en la IA no necesitan descansar y pueden analizar grandes volúmenes de información a la vez. Asimismo, la proporción de errores es significativamente menor en las máquinas que realizan las mismas tareas que sus contrapartes humanas. (Rouhiainen, 2018).

2.6 Operacionalización de Variables

Tabla 1.

Matriz de la Operacionalización de Variables

| VARIABLES | DEFINICIÓN CONCEPTUAL | DEFINICIÓN OPERACIONAL | DIMENSIONES | INDICADORES | ESCALA DE MEDICIÓN | ESCALA VALORATIVA |
|--|--|---|---|----------------------------|-----------------------|-----------------------|
| PREDICCIÓN DE AGENTES PATOLÓGICOS MEDIANTE PARÁMETROS DE LA CALIDAD DEL AGUA | El análisis predictivo se fundamenta en la recopilación e interpretación de datos previos para estimar tendencias y comportamientos futuros, cuya precisión depende de la dinámica y variabilidad del fenómeno estudiado. En estudios ambientales y sanitarios, este proceso resulta clave para evaluar la presencia de agentes patológicos microorganismos con capacidad de producir enfermedad y para determinar la calidad del agua, medida a través de parámetros físicos, químicos y biológicos que | La forma de medición se realizará mediante la recolección de registros de calidad del agua, donde los parámetros físico-químicos son condicionantes para la presencia de parámetros microbiológicos. Es por ello que, con el fin de ejecutar la simulación y predecir los parámetros microbiológicos son esenciales los parámetros físico químicos. | Parámetro Microbiológico y Parasitológico | Coliformes Totales | (NMP/100ml) | Cuantitativa de razón |
| | | | | Aceites y grasas | (mg/L) | |
| | | | Cloruros | (mg/L) | | |
| | | | Conductividad | (μ S/cm) | | |
| | | | Demanda Bioquímica de Oxígeno | (DBO ₅) (mg/L) | | |
| | | | Dureza | (mg/L) | | |
| | | | Demanda Química de Oxígeno | (mg/L) | Cuantitativa de razón | |
| | | | Nitratos | (mg/L) | | |
| | | | Nitritos | (mg/L) | | |
| | | | Potencial de Hidrógeno | (unidad de pH) | | |
| | | | Sólidos Disueltos Totales | (mg/L) | | |
| | | | Sulfatos | (mg/L) | | |
| | | | Temperatura | (°C) | | |
| | | | | | | |
| | | Parámetros Físicos - Químicos | | | | |

| | | | | | | |
|---|---|---|---|--|---|------------------------------|
| | <p>establecen su idoneidad para la vida y el consumo humano. (Makridakis, 1997) (Murray, 2020) (Organización Mundial de la Salud, 2017)</p> | | | <p>Turbiedad</p> | <p>(UNT)</p> | |
| <p>APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL</p> | <p>La Inteligencia Artificial (IA) es la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano. Sin embargo, a diferencia de las personas, los dispositivos basados en la IA no necesitan descansar y pueden analizar grandes volúmenes de información a la vez. Asimismo, la proporción de errores es significativamente menor en las máquinas que realizan las mismas tareas que sus contrapartes humanas. (Lasse Petteri, 2018)</p> | <p>Trabajar el modelo para la predicción de parámetros microbiológicos a partir de parámetros físicos y químicos, haciendo uso de redes neuronales y basándose en datos obtenidos de investigaciones realizadas en laboratorio.</p> | <p>Machine Learning Regression (Regresión en el Aprendizaje Automático)</p> <hr/> <p>Neural Network (Redes Neuronales Artificiales)</p> | <p>Decision tree</p> <hr/> <p>Random forest</p> <hr/> <p>ANN (1000 neuronales iniciales y cuatro capas)</p> <hr/> <p>ANN (1000 neuronales iniciales y siete capas)</p> | <p>MSE RMSE SMAPE R²</p> <hr/> <p>MSE RMSE SMAPE R²</p> <hr/> <p>MSE RMSE MAPE SMAPE R²</p> <hr/> <p>MSE RMSE MAPE SMAPE R²</p> | <p>Cuantitativa de razón</p> |

Fuente: Elaboración propia

III. Marco Teórico

3.1 Antecedentes de la investigación

3.1.1 A nivel internacional

(Carrillo & Hilaño, 2023) desarrollo el trabajo de investigación titulado “Evaluación de la calidad de agua a través de Macroinvertebrados Bentónicos en la cuenca del río Kushapukos en el cantón Tiwintza, provincia de Moronasa Santiago 2022”. El objetivo del autor busca "determinar la calidad del agua a partir de la abundancia de macroinvertebrados acuáticos en cuatro áreas del río Kushapukos para un diagnóstico del estado actual del área en estudio". con respeto al método, se utilizó el método experimental cuantitativa para recolectar macroinvertebrados, se utilizaron 4 estaciones de muestreo en diferentes áreas de la cuenca; cada estación tenía un tramo de 45 a 50 metros de longitud., como técnica no se tomó ninguna; finalmente, la investigación concluyó que se registraron 8 géneros en 8 familias con 6 órdenes representativas, incluyendo Hemíptera, Ephemeroptera, Coleóptera, Trichoptera, Díptera y Odonata. El índice BMWP/Col utilizado para categorizar el área de estudio alcanzó un valor de 45 BMWP/Col, lo que corresponde a valores de clase III con una calidad "dudosa" en comparación con el valor 36-60 del índice que representa “Aguas moderadamente contaminadas”.

(Narvárez & Tipán, 2022) desarrollaron el trabajo de investigación titulado “Análisis de la calidad de agua y presencia de Microplásticos en la parte alta, media y baja del río Pumacunchi en la provincia de Cotopaxi, periodo 2022”. Los autores se plantean como objetivo “evaluar la calidad del agua en puntos estratégicos del río Pumacunchi de la provincia de Cotopaxi, caracterizando el recurso hídrico con relación a la presencia de contaminantes y microplásticos".; el método fue cualitativo e inductivo; la muestra fue en

tres puntos estratégicos (P1 Cochapamba, P2 Canchagua y P3 Chantilin) del río Pumacunchi, el resultado fue que la calidad del agua presenta un índice de 82 en P1, lo que indica una calidad buena. Los parámetros que sobrepasaron los límites máximos permisibles (LMP) fueron DBO ($>5\text{mg/L}$) y oxígeno disuelto ($>4\text{mg/L}$). Los parámetros que superan los LMP son DBO ($>5\text{mg/L}$) y oxígeno disuelto ($>4\text{mg/L}$), y el P2 presenta un índice de 86 que indica una calidad de agua buena. Los parámetros que sobrepasan los LMP son: DBO ($>5\text{mg/L}$), oxígeno disuelto ($>5\text{mg/L}$), CE ($>750\mu\text{S/cm}$), SDT ($>500\text{mg/L}$) y arsénico ($>0,1\text{mg/L}$). El P3 presenta un índice de 67 que indica una calidad de agua regular. Por lo que se concluyó que los resultados de los dos muestreos de microplásticos fueron los siguientes: en el primer muestreo, P1 y P2, se encontraron microplásticos con morfología tipo fibra que representaban el 100% del análisis; en el segundo muestreo, P3, se encontraron microplásticos con morfología tipo fragmentos y morfología tipo fibra, con el 89 % correspondiendo a fibra y el 11 % a fragmentos de microplásticos.

(Salazar, 2020) desarrollo el trabajo de investigación titulado “Análisis de la calidad de agua del río Jerusalén afluente del río Pedro Carbo – Guayas, para uso agrícola”. El autor se plantea como objetivo “Analizar la calidad de agua del afluente de río Pedro Carbo, provincia del Guayas para uso agrícola”; se utilizó dos métodos, métodos teóricos (Inductivo y deductivo) y método científico empírico; como muestra se seleccionó uno cada quince días de los tres lugares donde los residentes reciben agua del río Jerusalén, afluente del río Pedro Carbo, el resultado de la investigación fue que la conductividad eléctrica supera los límites permitidos de $0.0007\ (\mu\text{S/cm})$, el oxígeno disuelto supera los límites permitidos de $6\ \text{mg/L}$, el nitrógeno total supera los límites permitidos de $40\ \text{mg/L}$ y la demanda bioquímica de oxígeno supera los límites permitidos de $2\ \text{mg/L}$, según los criterios

de calidad de aguas para uso agrícola o de riego, por lo que se concluye que el índice de calidad de agua ayuda a conocer el grado de contaminación que posee y los distintos tratamientos que se debe realizar al agua, para los diferentes usos consuntivos y no consuntivos que se pueden destinar el agua del río Jerusalén afluente del río Pedro Carbo.

(Maiquiza & Tonato, 2020) desarrollaron el trabajo de investigación titulado “Identificación de Diatomeas Epilíticas asociadas a la calidad del agua en el río Yanayacu, sector San Juan, Cantón Salcedo, provincia de Cotopaxi, 2020”. Los autores se plantean como objetivo “Identificar diatomeas epilíticas relacionadas con la calidad del agua del sector San Juan del río Yanayacu en el cantón Salcedo, provincia de Cotopaxi, 2020”; se aplicó el método analítico no experimental; las muestras se recolectaron en febrero y se seleccionaron 1 a 3 piedras de 15 a 20 cm. Los resultados se valoraron a través del programa IQADData, donde en el punto alto se obtuvo un resultado de 53,37 considerado una calificación Regular, punto medio un resultado de 41,75 considerado una calificación Mala y en el punto bajo 35,7 considerado una calificación Mala, posteriormente se comparó el índice trófico de calidad del agua (ITCA) teórico y de tablas; en el punto medio el ITCA teórico obtuvo un valor de 2,0 que comparado con el ITCA de las tablas se encuentra en el rango de 1,5 a 2,5 mostrando un nivel de contaminación moderada β -mesotrófico. Por lo que se concluye que tenemos identificado 26 especies de diatomeas epilíticas en el río.

(Espinoza, 2020) desarrolló el trabajo de investigación titulado “Evaluación de la calidad de agua y metales traza del río Jubones en el cantón Pasaje, provincia de El Oro, Ecuador”. El autor se plantea como objetivo “Evaluar la calidad del agua y la presencia de metales traza en el río Jubones, cantón Pasaje, Provincia de El Oro, utilizando indicadores fisicoquímicos”; Se empleó la técnica "área-velocidad o método de flotación"; con respecto

a la muestra, Cada punto de muestreo recibió una muestra de sedimento y una de agua, en total tres muestras de sedimento y tres muestras de agua.; de acuerdo con los resultados del cálculo del ICA, el recurso en los ítems P1, P2 y P3 se encuentra en el rango de calidad MEDIA o REGULAR según la clasificación de Brown. Además, los parámetros de DQO, conductividad y sólidos suspendidos totales se analizaron. dando como resultado que los coliformes fecales, exceden considerablemente los límites máximos permisibles. finalmente, La investigación llegó a la conclusión de que se determinó la presencia de metales traza (Plomo, Cadmio, Arsénico y Mercurio) en los sedimentos del río Jubones.

3.1.2 A nivel nacional

(Marin y Marin, 2022) desarrollaron el trabajo de investigación titulado “Impacto de las Fuentes de Contaminación Ambiental en la Calidad de Agua del Río Llaucano, Bambamarca, Cajamarca, 2021”. Los autores se plantean como objetivo “Determinar el impacto de las fuentes de contaminación ambiental en la calidad de agua del río Llaucano, Bambamarca, Cajamarca, durante el año 2021”; se aplicó el método científico con enfoque aplicativo; la muestra fue seis estaciones de muestreo para determinar las fuentes que afectan el río de manera intermitente e intermitente.; como resultado existió impacto directo entre la prevalencia de las fuentes de contaminación ambiental, en cuanto a la prevalencia de contaminantes en el cuerpo de agua analizado, habiendo contado con una sigma inferior a 0.050 que ha validado ello, por lo que se concluye que la fuente que ha generado una mayor contaminación ha estado relacionada con -307, siendo seguida por fuentes industriales, en donde se ha contado con un valor de contaminación de -135.

(Cuya y García, 2022) desarrollaron el trabajo de investigación titulado “Evaluación de la Calidad Ambiental del Agua Potable y Percepción en la Comunidad de Huaychao,

Acos Vinchos - Huamanga- Ayacucho – 2022”. Los autores se plantean como objetivo “Identificar las diferencias descriptivas entre la percepción de la comunidad sobre la calidad del agua para consumo en Huaychao _ distrito de Acos Vinchos _ Ayacucho”; con respecto al método, se aplicó un enfoque cuantitativo, de método inductivo de tipo básica de diseño no experimental descriptivo, transversal y correlaciona; la muestra se llevó a cabo a través de encuestas a 132 hogares, pero el investigador la amplió a 200 usuarios encuestados, como técnica se aplicó la recolección de datos denominada “encuesta”. Los resultados evidencian: existe un 20% de elementos analizados en la muestra del Reservorio que, exceden el límite máximo permisible establecido por la normatividad “Decreto Supremo N° 031-2010-SA - Reglamento de la Calidad del Agua para Consumo Humano” lo cual refleja características de agua para consumo de baja calidad. Concluyéndose que existen diferencias entre las muestras analizadas como indicadores de calidad en donde la quinta parte de elementos analizados presentan serios riesgos para la salud.

(Vargas, 2021) desarrollo el trabajo de investigación titulado “Determinación de índice simplificado de calidad de agua en el río Chili, Arequipa 2019”. El autor se plantea como objetivo “Determinar el Índice de Calidad de Agua Simplificado del Río Chili”; con respecto al método, se utilizó el método experimental cuantitativa; en la muestra se consideraron cinco puntos de muestreo, teniendo como resultado que los puntos de muestreo RIO 02, RIO 03 y RIO 04 superan el estándar de calidad ambiental que ningún parámetro físico químico supera el estándar de calidad ambiental; respecto al índice simplificado de calidad de agua. Por lo que se concluye que todos los puntos de muestreo se encuentran en el rango de 71 a 90 de la escala arbitraria el mismo que corresponde a una calidad buena de agua; sin embargo, el índice microbiológico establecido por la presencia de coliformes

termotolerantes que se encuentran por encima del estándar de calidad ambiental indican que los puntos de muestreo RIO 02, RIO 03 y RIO 04 presentan una mala calidad de agua.

(Mamani, 2017) desarrollaron el trabajo de investigación titulado “Calidad del agua del río Cumbaza respecto al efluente de la planta de tratamiento, San Roque de Cumbaza, 2021”. Los autores se plantean como objetivo “Evaluación de la calidad del agua del río Cumbaza en comparación con el efluente de la PTAR de San Roque de Cumbaza”; se aplicó el método aplicativo con enfoque cuantitativo; La muestra se realizó en 5 pasos : primer muestreo de agua del rio y efluente, remoción superficial del material sedimentado en laguna de desinfección, pruebas jarras para escoger floculante y/o coadyuvante, aplicación del floculante a laguna desinfección y segundo muestreo agua del rio y efluente. Los resultados del primer monitoreo, para los parámetros DQO, AyG, CT, pH, SST y T° encontraron dentro o por debajo de la normativa y el DBO del efluente superó por 5,9 mg/L del límite máximo establecidos. La remoción de los lodos se realizó superficialmente, mediante las pruebas de jarras se escogió la dosis 45 mg/L del floculante sulfato de aluminio y después del segundo monitoreo, los siete parámetros evaluados se encuentran por debajo o en los rangos establecidos por la normativa, por lo que se concluye que flocular el agua laguna de desinfección con sulfato de aluminio, disminuyo en el efluente los coliformes termotolerantes de 9,200 a 2,400 NMP/100 mL, siendo importante minimizar la presencia de solidos suspendidos y microalgas, ya que permite a los rayos solares realizar el proceso de desinfección.

(Vargas, 2020) desarrollo el trabajo de investigación titulado “Calidad de agua por parámetros fisicoquímicos y microbiológicos, rio Shilcayo sector Chontamuyo para proponer estrategias de conservación, San Martin – 2020”. El autor se plantean como

objetivo “Determinar la calidad del agua del sector Chontamuyo del río Shilcayo utilizando datos fisicoquímicos y microbiológicos para sugerir medidas de conservación”, se aplicó el método aplicativo con enfoque cuantitativo ; la muestra se tomó en 10 puntos de muestreo a lo largo de 2.5 km en el sector Chontamuyo; teniendo como resultado que los parámetros físicos fueron; el pH oscila entre 8.1- 8.6; la temperatura oscila entre 25.0-26.5 °C; Caudal promedio: 1.54 m³/s; SST promedio : 41 mg/L y los valores de Turbiedad promedio: 38,8 UNT; los parámetros químicos fueron ; DQO promedio : 104,9 mg/L y los valores de la DBO promedio:75.6 mg/L y los parámetros microbiológicos fueron; Coliformes Termotolerantes promedio: 86.890137,8 NPM/100mL y los resultados de la encuesta fueron el 95% de la población encuestada; finalmente, la investigación concluye que la calidad del agua para pH, Temperatura y SST se califica de buena y para los parámetros como DBO y Coliformes termotolerantes se califica de mala, por lo que se propone manejo de conservación de la zona en estudio y se recomienda a la municipalidad trabajar de manera articulada con la población.

3.1.3 A nivel regional y local

(Ramos, 2022) desarrollo el trabajo de investigación titulado “Formulación de índices de contaminación del agua (Ico) para un río altoandino: Caso subcuenca Chumbao, Andahuaylas, periodo 2018-2019”. El autor se plantea como objetivo “Establecer índices de contaminación de agua para el caso del río altoandino Subcuenca Chumbao”; como técnica se tomó la recolección y procesamiento de datos; para la muestra fue tomada la información de los parámetros del agua de la subcuenca Chumbao, correspondientes al periodo de monitoreo 2018–2019; teniendo como resultado que los parámetros con mayor incidencia en la calidad del agua fueron color, conductividad, OD, DBO, amoníaco, fósforo,

plomo, cromo y coliformes termotolerantes reportándose “baja” a “muy alta” contaminación; finalmente, se concluye que los índices de contaminación propuestos son instrumentos de gestión hídrica que permiten evaluar la calidad del agua.

(Solano, 2020) desarrollo el trabajo de investigación titulado “Monitoreo del índice de calidad de agua y su relación con la cafeína como contaminante emergente en la subcuenca del río Chumbao, Andahuaylas, Apurímac, Perú”. El autor se plantea como objetivo “Evaluar la relación entre la calidad del agua y el nivel de cafeína como contaminante emergente durante el monitoreo temporal y longitudinal en la subcuenca del río Chumbao de la cuenca Pampas, Andahuaylas, Apurímac, Perú”; con respecto al método, se aplicó el método científico con enfoque aplicativo; la muestra abarca una cuenca de 40.5 kilómetros y 143 kilómetros cuadrados con una altitud de 4100 a 2550 metros de altitud; como resultado se determinó el nivel de cafeína por cromatografía, los análisis se realizaron por triplicado, y los datos fueron analizados a través de un ANOVA bifactorial y comparación de media de Tukey al 5% de significancia. Se notó que el uso mayoritario de agua en la subcuenca del río Chumbao se dirige hacia la agricultura limitada, las zonas urbanas y los pastizales; los indicadores microbiológicos, el nivel trófico y el nivel de oxígeno aumentan a lo largo de todas las temporadas aguas abajo, mostrando una diferencia considerable entre puntos de muestreo y épocas ($p\text{-value} < 0.05$). Finalmente, la investigación concluye que los niveles de cafeína aumentan a lo largo del río, lo que indica que las zonas urbanizadas tienen niveles más altos. También se observaron diferencias significativas entre temporadas ($p\text{-value} < 0.05$), y en la temporada de estiaje se presentó una concentración más alta, con niveles de hasta 25.64 ± 0.40 ug/L.

3.2 Bases teóricas

3.2.1 *El agua*

La molécula del agua está formada por dos átomos de hidrógeno y uno de oxígeno (H_2O). En general, el término "agua" se refiere a la sustancia en su estado líquido, aunque también puede estar presente en su forma sólida, llamada hielo, y en su forma gaseosa, llamada vapor. Se encuentra principalmente en forma de vapor o hielo y es una sustancia bastante común en el universo y el sistema solar. Es fundamental para la supervivencia de todas las formas de vida conocidas. (Campbell & Reece, 2007)

3.2.2 *Calidad de agua*

El término "calidad del agua" se refiere a las propiedades químicas, físicas, biológicas y radiológicas del agua. (Diersing, 2009)

Es una medida de la calidad del agua en relación con las necesidades de una o más especies bióticas, así como con cualquier necesidad o propósito humano. (Johnson y Bassett, 1997)

Se utiliza con mayor frecuencia para referirse a un conjunto de reglas que permiten evaluar el cumplimiento. La salud de los ecosistemas, la seguridad de contacto humano y el agua potable son los estándares más comunes utilizados para evaluar la calidad del agua. (Diersing, 2009)

3.2.2.1 El valor de la calidad del agua. El aumento de la población, el aumento del consumo per cápita, la contaminación general de las fuentes de agua y el manejo inadecuado de las cuencas hidrográficas han reducido la disponibilidad de agua para consumo humano (Radulovich, 1997).

Es posible que ciertos elementos solubles se incorporen al agua y se vuelvan aún más peligrosos si estos elementos están en contacto directo con estas fuentes de agua, lo que resultará en enfermedades que afectan a la salud pública. Las consecuencias de consumir agua contaminada son múltiples: Según los expertos en salud pública, la ingestión del agua contaminada es la principal causa de aproximadamente un 80% de todas las enfermedades y más de una tercera parte de las defunciones en los países en vías de desarrollo. Se estima que la contaminación de agua por heces fecales afecta principalmente al 70% de la población que vive en áreas rurales de países en desarrollo. (OMS, 1998)

La escorrentía superficial, una forma de contaminación difusa o no localizada, está estrechamente relacionada con lo anterior. La contaminación por fuentes no localizadas es un factor importante en el aumento de los niveles de agentes patógenos en las fuentes de aguas superficiales, particularmente por coliformes fecales humanos y animales. En este sentido, un suministro seguro de agua para uso potable en cantidad, calidad y continuidad reduce la probabilidad de enfermedades transmitidas por la vía fecal y oral. (OMS, 1998)

3.2.3 La contaminación del agua

La contaminación hídrica, también conocida como contaminación del agua, es una alteración de esta que generalmente es causada por el ser humano, lo que lo hace inadecuado o peligroso para el consumo humano, la industria, la agricultura, la pesca y las actividades recreativas, así como para los animales. (Carta del Agua, 1968)

3.2.3.1 Tipos de contaminantes del agua

a. Contaminación acuática

Definición de estado del agua que ha cambiado sus condiciones naturales, físicas, químicas o bacteriológicas, necesarias para el proceso para el cual se le asignó inicialmente. Por ejemplo, un agua que no se considera contaminada para el uso de una caldera puede considerarse apropiada para el consumo humano o animal. (Becerril, 2009)

b. Contaminantes químicos antropogénicos

Los productos químicos que se encuentran en el agua incluyen una amplia gama de productos químicos antropogénicos, que son aquellos que se sintetizan o se producen indirectamente por la actividad humana. Los subproductos de la desinfección del agua potable, por ejemplo, y los "productos naturales", que son aquellas creadas por procesos fisicoquímicos o biológicos naturales. El geosmin, un alcohol bicíclico que da un sabor desagradable al agua y que es producido por algunas algas, es un ejemplo. Las emisiones industriales, la ineficiente disposición de residuos sólidos (como la incineración y los tiraderos de basura), los derrames accidentales, la aplicación controlada (como pesticidas en la agricultura, recarga artificial del agua subterránea, disposición en vertederos, etc.) y los contaminantes antropogénicos se han dispersado ampliamente en el ambiente y están "emergiendo" en las aguas superficiales y subterráneas. Aunque se han identificado todas estas fuentes, la importancia recae en las últimas, que incluyen rutas potenciales e importantes para eliminar el contaminante. Se ha descubierto que las actividades de consumo son una fuente potencialmente significativa de contaminación dispersa no controlada. (Becerril, 2009)

c. Contaminantes emergentes

Un gran porcentaje de los casos de contaminantes no regulados son "contaminantes emergentes", que pueden ser candidatos a regulación en el futuro, dependiendo de investigaciones sobre sus efectos significativos en la salud y datos de monitoreo sobre su incidencia. Los surfactantes, los productos farmacéuticos, los productos de cuidado personal, los aditivos de gasolina, los retardantes de fuego, los antisépticos, los aditivos industriales, los esteroides y las hormonas y los subproductos de desinfección del agua son ejemplos de compuestos que han emergido como particularmente relevantes recientemente. La característica de estos grupos de contaminantes es que, debido a sus altas tasas de transformación y eliminación, no necesitan permanecer en el ambiente para causar efectos negativos. La incidencia, la contribución de riesgo y los datos ecotoxicológicos no están disponibles para la mayoría de estos contaminantes emergentes. Por lo tanto, es difícil predecir qué consecuencias pueden tener para la salud de seres humanos y organismos acuáticos. (Barcelo, 2003)

3.2.4 Características físico-químico del agua.

3.2.4.1 Aceites y Grasas. Son sustancias lipídicas que al mezclarse con el agua permanecerán en la superficie, generando natas y espumas. Debido a que estas espumas y natas dificultan cualquier tratamiento físico o químico, deben eliminarse en los primeros pasos del tratamiento de un agua residual (Bernardina, 1986).

3.2.4.2 Cloruros. Son una de las sales que se hallan en los orígenes de suministros de agua y de desagüe (Muñoz, 2008).

Estos cloruros dan un sabor salado al agua; cuando se encuentra en forma de cloruro de sodio, el sabor se puede sentir a una concentración de 250 ppm de cloruro de sodio; esto se estableció principalmente por el sabor. (Muñoz, 2008)

3.2.4.3 Conductividad. La capacidad de una sustancia para conducir corriente eléctrica, o conductividad eléctrica, es lo contrario de la resistencia eléctrica. Es una variable que está influenciada por la cantidad de sales disueltas en un líquido. Siemens/cm (S/cm), microsiemens/cm ($\mu\text{S/cm}$) o milisiemens/cm (mS/cm) son las unidades de medición más comunes (De la Fuente, 2013).

Según (De la Fuente, 2013) “el valor de la conductividad en soluciones acuosas es directamente proporcional a la concentración de sólidos disueltos. Por lo tanto, cuanto mayor sea dicha concentración, mayor será la conductividad, como se puede ver en los valores siguientes”:

- ✓ Agua pura: 0,055 $\mu\text{S/cm}$
- ✓ Agua destilada: 0,5 $\mu\text{S/cm}$
- ✓ Agua de montaña: 1,0 $\mu\text{S/cm}$
- ✓ Agua de uso doméstico: 500 – 800 $\mu\text{S/cm}$
- ✓ Agua de mar: 50.000 – 60.000 $\mu\text{S/cm}$

3.2.4.4 Demanda Bioquímica de Oxígeno (DBO). La demanda bioquímica de oxígeno, también conocida como DBO, es una medida que se utiliza para determinar la cantidad de contaminantes presentes en las aguas residuales o aguas de desecho tanto domésticas como industriales. El término se refiere a la cantidad de oxígeno que las bacterias necesitan para estabilizar la materia orgánica biodegradable presente en las aguas

de desecho bajo condiciones aerobias. La cantidad de oxígeno utilizada se expresa en mg/L de oxígeno. (Norduzian, 1984)

Según (Norduzian, 1984) “la fuente principal para el procedimiento estándar de análisis de agua y aguas residuales es Standard Methods (ref. 2), que se actualiza y se vuelve a estandarizar cada cinco años”. Según las técnicas estándares, la prueba DBOS estándar se lleva a cabo durante cinco días a 20 grados centígrados, lo cual se discutirá más adelante. DBO total, DBO soluble y DBO suspendida son los tres tipos de DBO que se pueden atribuir a las dos fracciones del sustrato porque el material orgánico presente en el agua de desecho se presenta principalmente en dos formas: soluble y suspendido. Es común no intentar medir el DBO suspendido porque es extremadamente difícil, si no imposible, medirlo directamente. La DBO total y la soluble de una muestra se calculan primero, luego se calcula la DBO suspendida, ya que:

$$DBO\ total = DBO\ soluble + DBO\ suspendida$$

Por tanto:

$$DBO\ suspendida = DBO\ total - DBO\ soluble$$

3.2.4.5 Dureza Los iones metálicos bivalentes pueden reaccionar con el detergente y crear precipitados, y los aniones en el agua pueden crear incrustaciones, lo que determina la dureza del agua (Blasco, 1998).

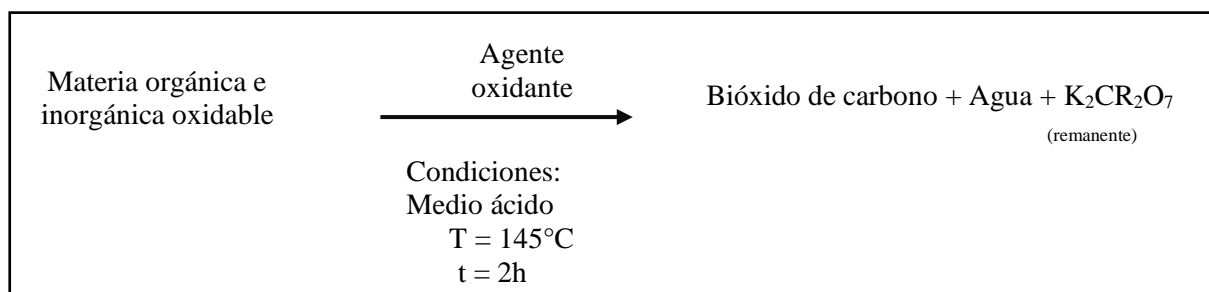
La dureza del agua varía según el lugar; las aguas superficiales son más blandas que las aguas subterráneas. La dureza del agua también revela el entorno de las formaciones físicas con las que se ha relacionado. (Blasco, 1998)

3.2.4.6 Demanda Química de Oxígeno (DQO). La demanda química de oxígeno (DQO) es un parámetro químico que se mide en miligramos de "oxígeno", equivalentes a la fracción orgánica disuelta y/o suspendida por litro de disolución (agua residual). La DQO es una medida de toda la materia orgánica e inorgánica presente en disolución y/o suspendida que puede ser químicamente oxidada por la acción de agentes oxidantes en condiciones ácidas (Blasco, 1998).

Como se puede observar en la Figura 1

Figura 1

Materia orgánica oxidada a bióxido de carbono y agua



Nota. Adaptado de Demanda Química De Oxígeno De Muestras Acuosas [Fotografía], por Mg. del Carmen Durán Domínguez de Bazúa, 2008.

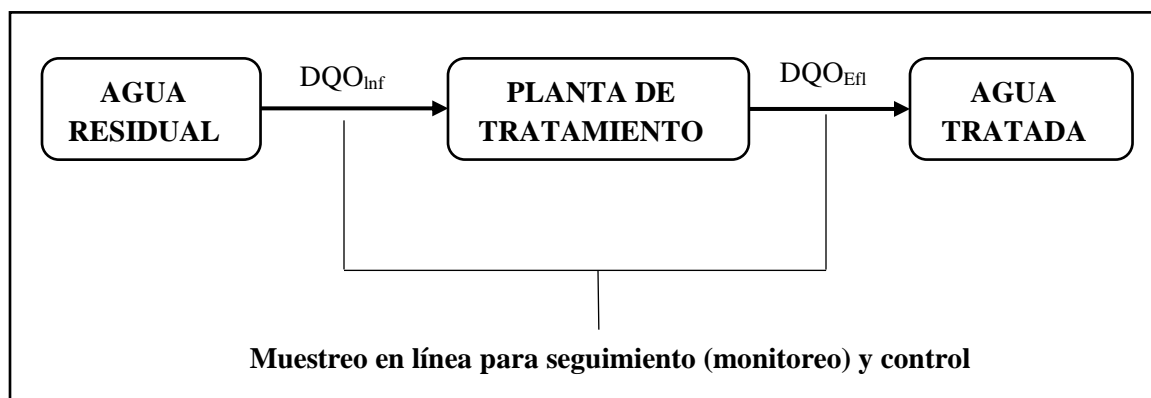
a. Importancia de la Demanda Química de Oxígeno.

Según (Ramírez y Duran, 2008) “la calidad de vida (DQO) es un indicador crucial y lo suficientemente rápido para determinar el nivel de contaminación del agua. También puede utilizarse para estimar la eficiencia de una planta de tratamiento de aguas residuales”.

Este parámetro permite la monitorización y el seguimiento en línea de la calidad del agua residual en tratamiento mediante la medición de la DQO_{Inf} (a la entrada) y la DQO_{Efl} (a la salida). (Ramirez y Duran, 2008) Como se puede observar en la Figura 2.

Figura 2

Seguimiento de una planta de tratamiento de aguas residuales



Nota. Adaptado de Demanda Química De Oxígeno De Muestras Acuosas [Fotografía], por Mg. del Carmen Durán Dominguez de Bazúa, 2008.

Se puede determinar si el tratamiento es adecuado y el porcentaje de eficiencia de depuración en el agua de salida mediante esta evaluación monitoria en línea. (Ramirez y Duran, 2008)

3.2.4.7 Nitratos. Debido a la polaridad del ión, los nitratos (sales del ácido nítrico, HNO₃) son muy solubles en agua. Su estructura es plana y de alta estabilidad, con nitrógeno en el centro y oxígeno en las esquinas de un triángulo equilátero. Cada uno de los enlaces N-O es un híbrido de un enlace sencillo y uno doble. En los sistemas acuáticos y terrestres,

los materiales nitrogenados tienden a transformarse en nitratos porque esta es la forma del nitrógeno más estable termodinámicamente en presencia de oxígeno (Albert y Lilia, 1997).

3.2.4.8 Nitritos. Los nitritos, también conocidos como sales del ácido nitroso, o HNO_2 , son solubles en agua y generalmente más estables que el ácido del que provienen. Se forman naturalmente a partir de nitratos a través de la reducción bacteriana o la oxidación incompleta del nitrógeno en los sistemas acuáticos y terrestres (Albert y Lilia, 1997).

El ión nitrito es menos estable que el ión nitrato porque es muy reactivo y puede funcionar como agente oxidante y reductor. Como resultado, solo se encuentran cantidades significativas en condiciones de baja oxigenación. Esto explica la rápida transformación de los nitritos en nitratos y la prevalencia de estos últimos en las aguas, tanto superficiales como subterráneas, así como en el ambiente. Los factores abióticos y los biológicos pueden causar esta reacción de oxidación. (Albert y Lilia, 1997)

3.2.4.9 Potencial de Hidrógeno (pH). El valor del pH es el que determina si una sustancia es ácida, neutra o básica calculando la cantidad de iones hidrogeno presentes en ella. La sustancia está clasificada como neutra en la escala de 0 a 14. Los valores de pH por debajo de 7 indican acidez y los valores por encima de 7 indican base. (Albert y Lilia, 1997)

Según (Albert y Lilia, 1997) menciona que:

a. Características

El pH de las aguas naturales está influenciado por la composición del suelo circundante: un pH alcalino indica que el suelo es calcáreo y un pH ácido indica que el suelo es silíceo. El pH de las aguas contaminadas en áreas de descargas industriales puede ser extremadamente ácido.

Según (Albert y Lilia, 1997) menciona que:

b. Riesgo

Un pH alto o bajo en aguas naturales puede alterar el equilibrio químico del agua y liberar contaminantes, lo que resulta en condiciones tóxicas. Los problemas pueden reducir las poblaciones de organismos acuáticos.

3.2.4.10 Sólidos Disueltos Totales. La cantidad total de sólidos disueltos presentes en el agua se denomina sólidos totales disueltos. Sustancias orgánicas e inorgánicas que se pueden resolver en agua. Todos los sólidos ionizados están en solución. Los sólidos en suspensión, los coloides y los gases disueltos no están incluidos. (Blasco, 1998)

a. Características

La conductividad eléctrica está estrechamente relacionada con los sólidos totales disueltos. La conductividad eléctrica aumenta con la cantidad de sales disueltas en el agua. Los iones disueltos son la mayoría de los sólidos que permanecen en el agua después de la filtración de arena. Por ejemplo, el agua contiene cloruro de sodio como Na^+ y Cl^- . El agua de alta pureza, que idealmente solo contiene H_2O sin sales o minerales, tiene una baja conductividad eléctrica. El contenido total de iones disueltos en el agua aumenta de un 2 a un 3 % por grado Celsius como resultado de la influencia de la temperatura del agua en la conductividad eléctrica. (Blasco, 1998)

Según (Blasco, 1998) menciona que:

b. Riesgo

El aumento del caudal de los ríos ha arrastrado materiales, lo que ha provocado una alta concentración de sólidos disueltos. Esta concentración es un factor importante en la determinación de los usos del agua.

3.2.4.11 Sulfatos. Los sulfatos son sales o ésteres del ácido sulfúrico. Hay un átomo de azufre en el centro de un tetraedro con cuatro átomos de oxígeno en los vértices. La mayoría de las industrias utilizan sales, ácidos derivados y peróxidos de sulfato (Don y Perry, 2008).

a. Los sulfatos inorgánicos y orgánicos

Las sales del ácido sulfúrico se llaman sulfatos inorgánicos. Se pueden encontrar en la naturaleza en forma de yeso o aljez (sulfato de calcio dihidratado $\text{CaSO}_4 \cdot 2 \text{H}_2\text{O}$), baritina (sulfato de bario) o sulfato de sodio (Na_2SO_4). Se forman a partir de sulfuros de otros elementos cuando entran en contacto con el oxígeno del aire por oxidación. (Don y Perry, 2008)

Los ésteres del ácido sulfúrico se denominan sulfatos orgánicos con la fórmula genérica $\text{R-O-SO}_2\text{-O-R'}$. Se pueden formar directamente a partir del alcohol correspondiente, del ácido sulfúrico, del trióxido de azufre o del cloruro de sulfurilo (Cl_2SO_2) en presencia de una base. (Don y Perry, 2008)

3.2.6.12 Temperatura. La energía cinética media de las moléculas de agua se mide por la temperatura. Se mide en grados Fahrenheit o grados Centígrados, respectivamente. (Agua Rio, 2017)

A. Importancia de la temperatura

Es uno de los criterios más cruciales para la calidad del agua. La química del agua y las funciones de los organismos acuáticos están influenciadas por la temperatura. (Agua Rio, 2017)

Según (Agua Rio, 2017), la temperatura tiene un impacto en:

- ✓ La cantidad de oxígeno disponible para disolverse en agua.
- ✓ La rapidez con la que se produce la fotosíntesis en plantas acuáticas como las algas.
- ✓ La tasa de metabolismo de los seres vivos.
- ✓ Los organismos son susceptibles a los desechos tóxicos, parásitos y enfermedades.
- ✓ Periodos de crecimiento, migración y estivación de los organismos acuáticos.

B. Factores que afectan la temperatura.

Los factores que afectan según (Agua Rio, 2017) son:

a) Factores naturales:

- ✓ **Energía solar:** Sombra, temperatura del aire, cambios estacionales y diarios el flujo.
- ✓ **Nivel de profundidad del agua:** La corriente de agua subterránea es generalmente más fría que la corriente de agua superficial. La cantidad de agua de la superficie que fluye a una temperatura diferente de la corriente.

b) Factores humanos:

- ✓ Eliminación de la flora riparia.
- ✓ Aceleración de la erosión de la tierra, aumento de la turbidez y absorción de calor
- ✓ Aumento de la escorrentía de aguas pluviales

- ✓ Cambios en la corriente morfológica (profundidad de los pozos), el substrato y el flujo
- ✓ Descargas de agua fría de las plantas de energía eléctrica.

3.2.4.13 Turbiedad. La turbidez, también conocida como turbiedad, es la falta de transparencia de un líquido como resultado de la presencia de partículas en suspensión. Cuanto mayor sea la cantidad de sólidos en suspensión presente en un líquido (normalmente llamado agua), más sucio se verá y aumentará la turbidez. La turbidez es una medida importante de la calidad del agua; cuanto más turbia es el agua, más baja es la calidad del agua (Perez y Mariscal, 1999).

La presencia de partículas en suspensión hace que el agua sea turbia. Las partículas pequeñas (de diámetro inferior al micrón) requieren tratamiento porque tienen una baja velocidad de sedimentación. (Perez y Mariscal, 1999)

Las partículas de tamaño superior a un micrón sedimentan naturalmente. Algunas son de naturaleza orgánica (bacterias, parásitos, algas, zooplancton, ácidos fúlvicos y coloides húmicos), mientras que otras son de naturaleza inorgánica (arcilla, fangos y óxidos minerales) que se derivan de la erosión del suelo. Además de las fuentes naturales, las actividades humanas producen efluentes cargados de estas partículas y otras sustancias que pueden combinarse con ellas (virus entéricos, contaminantes químicos, cloro, etc.) en el cuerpo de agua, las plantas y las redes de distribución. Los tratamientos efectivos para eliminarlos variarán según el tamaño, la composición química y otras variables dependientes del agua. (Perez y Mariscal, 1999)

3.2.4.14 Sólidos Suspendidos Totales. Según (Blasco, 1998) menciona que los “SST es un índice utilizado en el tratamiento de aguas residuales y la evaluación de la calidad del agua. Indica la cantidad total de sólidos presentes. Los sólidos que se componen de sólidos coloidales, sólidos sedimentables y sólidos en suspensión que tienen un tamaño de partícula que no supera el diámetro de un filtro de fibra de vidrio típico de 1.5 micrómetros o su equivalente”.

Según (Blasco, 1998) clasifica por:

a. Características

- ✓ Este indicador indica la carga total de sólidos suspendidos en cuerpos de agua en lugar de vertimientos. En términos técnicos, la carga indica el volumen de sólidos suspendidos que fluyen o se alojan en un cuerpo de agua durante un período de tiempo específico.
- ✓ La mayoría de las veces, las sustancias no disueltas se denominan materia suspendida o sólidos suspendidos. Pocas veces se realizan pruebas de sólidos suspendidos, y su evaluación generalmente se realiza mediante medición de turbiedad.
- ✓ Los residuos domésticos se evalúan utilizando sólidos suspendidos y sólidos suspendidos volátiles.

b. Riesgo

- ✓ Debido a la descomposición de los materiales volátiles en los sólidos, los altos niveles de sólidos suspendidos totales pueden ser dañinos para los hábitats bénticos y causar condiciones anaerobias en el lecho de los lagos, ríos y mares.

- ✓ Las partículas suspendidas en el agua facilitan la adhesión de metales pesados y otros materiales orgánicos tóxicos y pesticidas en el agua, lo que altera la calidad del agua con el fin de preservar el medio ambiente.
- ✓ Las aguas turbias se vuelven más calientes debido al calor absorbido por las partículas suspendidas, lo que reduce la concentración de oxígeno en el agua. El oxígeno se disuelve mejor en aguas más frías. Además, algunos organismos no son capaces de sobrevivir en aguas más cálidas.
- ✓ Las partículas en suspensión dispersan la luz, lo que reduce la actividad fotosintética en plantas y algas, lo que aumenta aún más la concentración de oxígeno.
- ✓ Los lagos poco profundos se colmatan más rápido debido a la sedimentación de partículas en el fondo, lo que hace que los huevos de peces y las larvas de insectos se cubran y sofoquen, y las agallas se tupen o se dañen.
- ✓ El plancton y los materiales suspendidos inorgánicos disminuyen la penetración de la luz en el cuerpo de agua, lo que reduce la producción primaria y, por lo tanto, la cantidad de alimento disponible para los peces.

3.2.5 Características microbiológicas del agua

Los coliformes son un grupo de bacterias que comparten una serie de cualidades. Todas las bacterias de la familia Enterobacteriaceae son comunes en el tracto digestivo de algunos animales. (Lopez, 2018)

Su nombre proviene de la bacteria *Escherichia coli*, la más conocida de este grupo. Sin embargo, las bacterias coliformes pertenecen a cuatro géneros diferentes: *Escherichia*, *Klebsiella*, *Enterobacter* y *Citrobacter*. (Lopez, 2018)

Debido a que estas bacterias son un indicador de los niveles de contaminación en los cuerpos de agua, son ecológicamente significativas. La existencia de estas bacterias indica que la materia fecal está presente en el agua. Por lo tanto, algunas son patógenos significativos para los humanos. (Lopez, 2018)

Según (Lopez, 2018); señala la clasificación taxonómica de las bacterias coliformes:

- ✓ **Dominio:** Bacteria
- ✓ **Reino:** Monera
- ✓ **Filo:** Proteobacteria
- ✓ **Clase:** Gammaproteobacteria
- ✓ **Orden:** Enterobacteriales
- ✓ **Familia:** Enterobacteriaceae.
- ✓ **Géneros:** *Escherichia*, *Enterobacter*, *Citrobacter*, *Kleibsellla* *Salmonella typhi*, *Shigellae dysenteriae*

Para satisfacer las necesidades de los humanos, el recurso hídrico distribuido debe estar libre de coliformes totales, coliformes termotolerantes, bacterias E. Coli, organismos libres y algas. (Minsa, 2010)

Según (Lopez, 2018) menciona que: “los organismos coliformes se dividen en dos grandes categorías: totales y termotolerantes”.

3.2.7.1 Coliformes totales. Las bacterias coliformes totales están formadas por una variedad de bacterias con forma bacilar, gramnegativas y anaerobios facultativos, lo que significa que pueden desarrollarse tanto con o sin oxígeno (Lopez, 2018).

Los coliformes totales incluyen todas las bacterias coliformes, algunas de las cuales son patógenas para los humanos y otras totalmente inofensivas. (Lopez, 2018)

Estas bacterias son un buen indicador de la limpieza o la contaminación de las aguas. Debido a que muchas de las bacterias coliformes se desarrollan y proliferan en la materia fecal, son un excelente indicador de la contaminación del agua por materia fecal. Mientras más coliformes hay en el agua, más contaminado está. (Lopez, 2018)

a. Fuente

Esto sugiere una contaminación y, por lo tanto, la presencia de coliformes, como lo demuestran las aguas negras, los vertidos domésticos, las granjas de animales y los lavaderos de carros. Debido a que no cumplen con los límites de calidad de agua más altos, estas aguas no son adecuadas para el consumo humano. (Minsa, 2010)

b. Características

Los coliformes totales son bacterias gram negativas con forma de bastones y crecen en presencia de sales biliares u otros agentes tensores activos. (Minsa, 2010)

Las bacterias del grupo coliforme viven en el intestino. El indicador bacteriano clave del recurso hídrico es E. Coli, que se encuentra principalmente en las heces de humanos y animales. (Minsa, 2010)

c. Riesgo

Los animales pueden experimentar pérdidas de animales y plantas, así como problemas gastrointestinales, epidemias y hepatitis A y B como resultado de la

contaminación del agua por coliformes fecales. Es necesario monitorear, vigilar y controlar la presencia de microorganismos. (Minsa, 2010)

Según (Lopez, 2018) “las bacterias coliformes incluyen unos géneros bien conocidos como”:

3.2.5.1.1 *Escherichia*. Gracias a su potencial patógeno, este género de bacterias es muy conocido y estudiado. La bacteria *Escherichia Coli* es la más característica de este grupo. Son bacterias de bacilo con bordes redondeados. De tamaño entre 0,5 y 3 micras, habitan en la flora intestinal de algunos animales como comensales (Lopez, 2018).

Las enfermedades causadas por las bacterias del género *Escherichia* se concentran principalmente en el sistema digestivo, como los síndromes gastrointestinales, cuyos síntomas principales son la diarrea y los cólicos. Por lo tanto, estas bacterias pueden causar problemas en el tracto genitourinario. (Lopez, 2018)

a. *Escherichia coli*. Es una bacteria que se adquiere ingiriendo alimentos y bebiendo agua contaminada, lo que aumenta la probabilidad de complicaciones en niños y ancianos. La diarrea, los vómitos y el dolor abdominal severo son los síntomas principales. Además, la infección puede provocar el síndrome hemolítico urémico (HUS) en niños menores de 5 años, que tiene problemas renales y destruye los glóbulos rojos. (Boston, 2019)

3.2.5.1.2 *Enterobacter*. En el campo médico, el género *Enterobacter* es bien conocido por causar una gran cantidad de infecciones. Se pueden encontrar principalmente en los suelos (algunos son descomponedores de materia orgánica muerta), en el agua y en algunos lácteos (Lopez, 2018).

3.2.5.1.3 *Klebsiella*. Estas bacterias viven principalmente en hábitats acuáticos y se reproducen rápidamente cuando tienen suficientes nutrientes (Lopez, 2018).

Este género tiene alrededor de 13 especies, pero las siguientes son las más importantes en términos de salud: *Klebsiella pneumoniae*, *Klebsiella ozaenae*, *Klebsiella oxytoca* y *Klebsiella rhinoscleromatis*. (Lopez, 2018)

Las infecciones de las vías respiratorias, las infecciones del tracto digestivo, las infecciones del tracto respiratorio superior y las infecciones de las vías respiratorias en general son causadas por estos agentes patógenos. (Lopez, 2018)

3.2.5.1.4 *Citrobacter*. Este es uno de los géneros de bacterias coliformes que causa una variedad de enfermedades en los humanos. Los abscesos cerebrales e infecciones del tracto urinario son ejemplos de estas afecciones (Lopez, 2018).

Son bacterias en forma de bacilo que pueden medir entre 2,5 y 6 micras. Son gram negativos, catalasa positivos y reducen nitratos a nitritos son algunas de las características que comparte con el resto de las bacterias coliformes. Además, es indol positivo y tiene la capacidad de fermentar la glucosa. (Lopez, 2018)

3.2.5.1.5 *Salmonella Typhi*. Es la bacteria que causa la fiebre tifoidea, que es causada por dolor de cabeza, fiebre, malestar, anorexia, dolor abdominal, diarrea o estreñimiento. Se obtiene al beber agua contaminada por las heces de una persona infectada (Khan y Shamim, 2022).

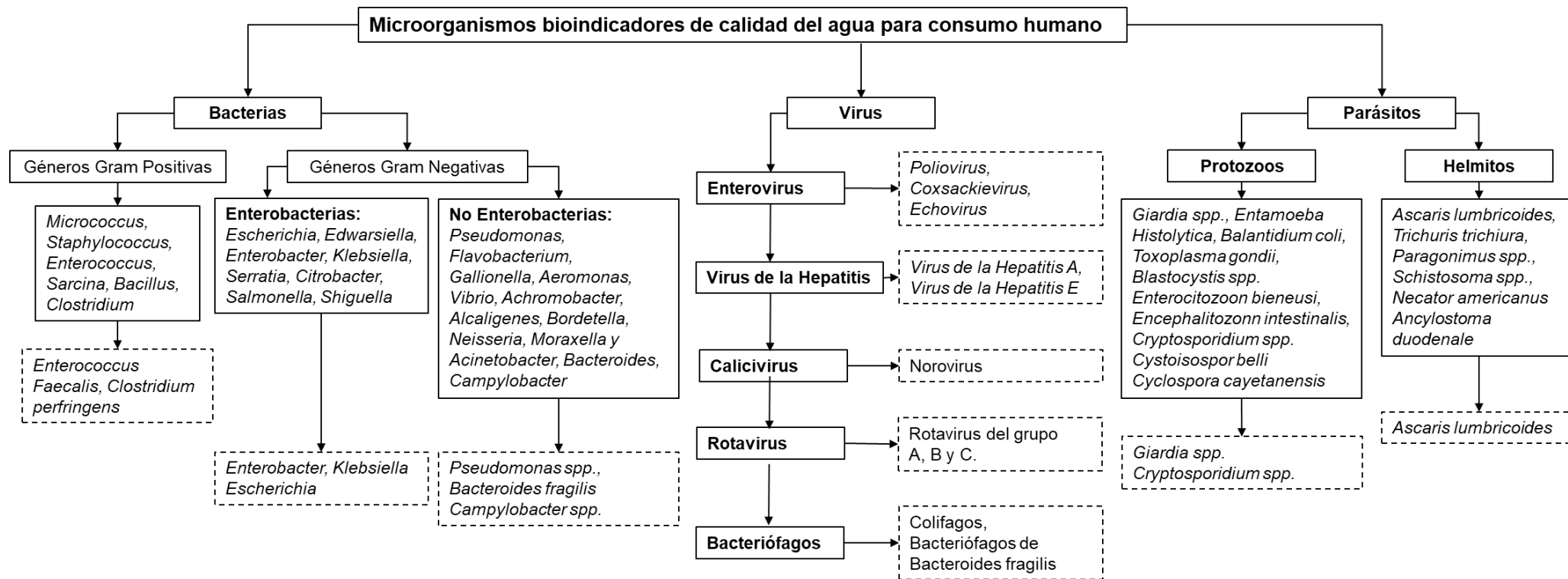
3.2.5.1.6 *Shigellae Dysenteriae*. Esta bacteria es responsable de causar fiebres altas, retorcijones, convulsiones, etc. Esta enfermedad puede causar epidemias masivas y altas tasas de mortalidad (Moncada y Luis, 2016).

3.2.5.2 Coliformes termotolerantes o fecales. Dentro de las bacterias coliformes totales, este es un subgrupo. Debido a su capacidad para fermentar la lactosa a temperaturas extremadamente altas (entre 44 y 45°C), se les conoce como termotolerantes (Lopez, 2018).

3.2.6 Microorganismos bioindicadores

Figura 3

Principales microorganismos bioindicadores



Nota. Adaptado de Patógenos e Indicadores Microbiológicos de Calidad del Agua para Consumo Humano [Fotografía], por Ríos Tobón, S.

Agudelo Cadavid, Ruth M. Gutiérrez Builes, Lina A, 2017, Rev. Fac. Nac. Salud Pública.

3.2.7 Factores que intervienen en la transmisión de enfermedades

Según (Guerra, 2001) indica que “en el agua existen varios agentes biológicos, como bacterias, virus y protozoos, que pueden ser dañinos o no para la salud de las personas”. Los agentes que no causan grandes problemas a la salud de las personas y pueden eliminarse fácilmente se denominan de riesgo potencial, mientras que los agentes que causan problemas serios a la salud de las personas se denominan riesgos reales. Además, es posible que los riesgos potenciales se conviertan en riesgos reales, lo que depende de una variedad de factores, como:

- ✓ La cantidad de excretas, la infección del huésped y el contacto del individuo con la dosis de infección son factores importantes.
- ✓ Existe alguna enfermedad

Según (Moscoso, 2016) “el tipo de características del agente patógeno, el riesgo a la salud de las personas puede aumentar”:

- ✓ Los patógenos se adhieren al entorno y su latencia o etapa de desarrollo es prolongada.
- ✓ La dosis de desinfección es baja y el sistema inmunológico del huésped es débil.
- ✓ Es posible que se propaguen de varias maneras, como por medio de agua, alimentos y malos hábitos de higiene personal.

3.2.8 Principales enfermedades por el consumo de aguas residuales

En 2015, alrededor del 71% de la población mundial utiliza un servicio seguro de agua potable que está disponible durante todo el día y no contamina. El resto de la población, que no tiene acceso a agua potable, se abastece de agua contaminada con excrementos de aguas residuales domésticas sin tratar, que desembocan en cuerpos de agua superficiales. (Wendorff, 2019)

En Perú, el 94% de las viviendas tienen acceso al agua potable en el ámbito urbano, mientras que el 59% en el ámbito rural, lo que indica que es importante priorizar el acceso a las poblaciones rurales. En cuanto al sistema de alcantarillado, la cobertura total en el país es del 76.8%, con un 90.2% en el ámbito urbano y solo el 29% en el ámbito rural. Además, solo el 78.6% de las aguas residuales de los desagües son enviadas a una planta de tratamiento de aguas residuales. (Wendorff, 2019)

El agua contaminada y el saneamiento ineficaz ponen en riesgo la salud de todos, causando una variedad de enfermedades, como las diarreas, que causan más de 842.000 muertes al año. (Valdes y Moreno, 2011)

Según (Valdes y Moreno, 2011) menciona; las enfermedades más comunes causadas por el consumo de aguas contaminadas son:

a. Cólera

Esta enfermedad se clasifica como portadora o transportada por el agua a través de la contaminación fecal. Actualmente, los brotes de cólera siguen ocurriendo en las comunidades rurales debido principalmente a un suministro inadecuado o ausente de agua potable y saneamiento. Además, los brotes de esta enfermedad también ocurren en lugares

donde las condiciones ambientales fundamentales han sido destruidas o afectadas por conflictos sociales o desastres naturales antrópicos, lo que aumenta el riesgo de una epidemia.

Desde hace mucho tiempo, el cólera ha estado presente en Haití. En 2010, comenzó la epidemia con 179 379 casos y una morbilidad del 56% de la población en solo 3 meses, resultando en un 2,2% de muertes. Esto se debe a que Haití es un país con bajos ingresos, lo que aumenta el riesgo de que se produzca una epidemia de este tipo. La principal razón de esto es que en el país hay una disponibilidad limitada de agua potable y saneamiento. (Sánchez y Cabezas, 2018).

En República Dominicana, durante el mismo año que ocurrió el brote de Haití, se registraron 31090 casos sospechosos de cólera y 458 fallecimientos. En Cuba se registraron 678 casos de cólera y 3 fallecimientos en 2013. En 1991, se produjo el primer brote de cólera en Perú, con 322 562 casos y alrededor de 3000 fallecimientos. (Miranda, 2013)

Además, no se han registrado casos de cólera en los últimos diez años, pero el acceso a agua potable y saneamiento es escaso, lo que aumenta el riesgo de una nueva epidemia. (Miranda, 2013) . En 1991, Perú experimentó un caso muy conocido de cólera que resultó en la muerte de 2909 personas y un total de 322 562 casos. Los primeros casos de diarrea se presentaron en Chimbote, Piura y Chancay debido a la falta de agua potable y problemas en los procesos de desinfección. Además, se detectaron presencia de coliformes fecales y bacterias en los procesos de desinfección. (Carcelèn y Moran, 2022)

La enfermedad afecta principalmente a las personas que viven en áreas rurales, a las comunidades de bajos recursos económicos que afectan a los niños o a las áreas afectadas

por el fenómeno del Niño, ya que la presencia de la enfermedad aumenta con la temperatura del agua. (Harvez y Àvila, 2013)

Por otro lado, si existe un tratamiento para el cólera y es mediante la aplicación de vacunas, En la actualidad, el conocimiento de esta enfermedad se difunde a nivel mundial a través de diferentes medios de comunicación o campañas, lo que facilita el acceso a las vacunas para toda la población, especialmente en las zonas rurales. (Miranda, 2013)

Según (Valdes y Moreno, 2011) menciona que:

b. Fiebre Tifoidea

Esta enfermedad también se incluye en las enfermedades transmitidas o transmitidas por el agua a través de la contaminación fecal. Debido a la falta de buenas condiciones de saneamiento y agua potable, esta enfermedad es más común en los países subdesarrollados. La OMS estimó que en 2018 había entre 11 y 20 millones de personas en todo el mundo que contraían esta enfermedad, con un promedio de 128 000 a 161 000 muertes.

En Perú, la fiebre tifoidea es común y se debe principalmente a problemas de saneamiento ambiental, lo que la convierte en una de las seis enfermedades más importantes del país. En zonas pobres, la incidencia es de 40 a 60 casos al año, con un promedio de 300 a 500 casos en adultos jóvenes, siendo el 35% niños menores de 14 años y el 65% adultos de 58 a 79 años. (Raymundo y Rafaela, 2018)

Los principales afectados son las comunidades pobres o las personas susceptibles a enfermedades. Además, esta enfermedad afecta principalmente a los niños de 5 a 15 años. En el tratamiento, se han mejorado los métodos de administración de vacunas según la edad.

Los niños mayores de dos años reciben vacunas inyectables y las vacunas en cápsulas se administran por vía oral a los niños mayores de cinco años. (Minaya y Rodríguez, 2018)

c. Hepatitis A

Este tipo de enfermedad se presenta principalmente por falta de higiene o infraestructuras sanitarias deficientes; la enfermedad puede llegar a ser de moderada a grave dependiendo de la edad del infectado. La hepatitis A se presenta de manera esporádica a nivel mundial y puede durar muchos meses en las personas afectadas. Además, la enfermedad puede causar problemas económicos y sociales en las comunidades afectadas porque la recuperación puede tardar meses. (Miranda, 2013)

d. Diarrea aguda

Debido a que se propaga de manera repentina y contagiosa, este tipo de enfermedad es común (Sánchez y Cabezas, 2018). Se estima que las enfermedades diarreicas agudas (EDAS) causan la muerte de 760 000 millones de niños cada año, siendo la segunda enfermedad con mayor mortalidad en niños menores de cinco años (Miranda, 2013)

En el año 2017, se registraron 67086 casos de diarrea aguda en Perú, incluyendo el 29% en menores de cuatro años y el 11% en menores de un año. Incluso se ha demostrado que la mayoría de los casos de EDAS ocurren en áreas rurales (Miranda, 2013)

La falta de agua potable, las deficiencias del sistema de saneamiento, los niños menores de un año, las áreas de pobreza extrema y la falta de higiene personal son las poblaciones más vulnerables (Ministerio de Salud, 2023)

Las diarreas agudas no tienen tratamiento, ya que pueden ser un síntoma de otras enfermedades. Sin embargo, el tratamiento para la deshidratación, como la aplicación de suero, existe (Bernaola y Chávez, 2011)

e. Infección urinaria

Es la infección más común causada por bacterias coliformes. Como estas bacterias son comunes en el tracto digestivo, es común que se muevan desde el ano hacia la uretra. Dado que la distancia entre el ano y el orificio uretral es menor que en el hombre, esto ocurre con mayor frecuencia en las mujeres. (Lopez, 2018)

Según (Lopez, 2018) menciona que: Las infecciones urinarias pueden presentar los siguientes síntomas:

- ✓ La presión en la pelvis
- ✓ Aumento de la frecuencia urinaria y las molestias en el abdomen inferior
- ✓ Ardor al orinar
- ✓ La sensación es que la vejiga no se ha vaciado completamente.
- ✓ Al orinar, hay dolor y ocasionalmente hay sangre en la orina.

La infección puede propagarse desde las vías urinarias inferiores hacia los riñones si estos síntomas no son tratados a tiempo. Si esto sucede, puede experimentar los siguientes síntomas y signos:

- ✓ Náuseas y vómitos, fiebre elevada y dolor en la espalda y los riñones.
- ✓ Escalofríos.

De lo contrario, estos síntomas pueden causar problemas clínicos más graves, como la insuficiencia renal, por lo que es crucial abordarlos tan pronto como aparecen.

Según (Lopez, 2018) menciona que:

f. Infección entérica

Son infecciones en el tracto gastrointestinal. Las bacterias coliformes son comunes responsables de este tipo de infecciones porque forman parte de la flora bacteriana del tubo digestivo.

Las infecciones de este tipo suelen presentar los siguientes síntomas:

- ✓ Frecuentes deposiciones que van desde pastosa hasta líquida
- ✓ Dolor en el abdomen
- ✓ Heces que contienen sangre y/o moco
- ✓ Vómitos, fiebre, cansancio y malestar general

g. Neumonía

Esta es una infección que tiene un impacto en los tejidos pulmonares. Algunas bacterias del género *Klebsiella* pueden causar neumonía. Puede ser mortal si no se trata a tiempo. (Lopez, 2018)

Los síntomas de neumonía incluyen:

- ✓ La tos puede presentar flema purulenta o sangrienta
- ✓ Fiebre alta
- ✓ Ansiedad y cansancio general
- ✓ Dolor en el pecho que puede empeorar al moverse

- ✓ Náuseas y gases
- ✓ Aumento de la presión arterial y aumento de la frecuencia cardíaca

h. Otras infecciones

Según (Lopez, 2018) “en situaciones extremas, las bacterias de la familia de las coliformes pueden causar infecciones adicionales a las mencionadas anteriormente”.

Estos incluyen:

- ✓ La rinitis atrófica, la endoftalmitis y la osteomielitis.
- ✓ Existen otitis media y celulitis

Un porcentaje muy pequeño de la población tiene este tipo de infecciones, especialmente aquellos con problemas en su sistema inmunológico que los hacen más susceptibles.

3.2.9 Inteligencia Artificial

La capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano se conoce como inteligencia artificial. Los dispositivos basados en la IA, por otro lado, pueden analizar simultáneamente una gran cantidad de datos, lo que los distingue de las personas. Además, la tasa de errores es significativamente menor en las máquinas que realizan tareas similares a las humanas. (Rouhiainen, 2018)

La idea de que las computadoras o los programas informáticos pueden tanto aprender como tomar decisiones es particularmente importante y algo de lo que deberíamos estar conscientes porque sus procesos están creciendo exponencialmente con el tiempo.

Estas dos habilidades han permitido que los sistemas de inteligencia artificial lleven a cabo una gran cantidad de tareas que antes solo eran posibles para los humanos. (Rouhiainen, 2018)

Ya se están utilizando tecnologías basadas en la IA para que los humanos disfruten de mejoras significativas y una mayor eficiencia en casi todos los ámbitos de la vida. Sin embargo, el rápido desarrollo de la IA nos obliga a ser cautelosos para evitar y evaluar posibles desventajas directas o indirectas que pueda generar su expansión. (Rouhiainen, 2018)

Según (Rouhiainen, 2018) indica que, la IA funciona en casi todas las situaciones. Estos son solo algunos de los numerosos usos técnicos de la inteligencia artificial que están creciendo rápidamente en la actualidad:

- ✓ El reconocimiento de imágenes estáticas, la clasificación y el etiquetado mejoraron el rendimiento de la táctica algorítmica comercial.
- ✓ Procesamiento de datos de pacientes eficiente y escalable.
- ✓ Mantener la predicción, detectar y clasificar objetos y difundir contenido en las redes sociales.
- ✓ Protección contra amenazas cibernéticas.

3.2.10 Aprendizaje automático (Machine Learning)

Uno de los enfoques principales de la inteligencia artificial es el aprendizaje automático. En pocas palabras, se trata de una rama de la informática en la que las computadoras o las máquinas pueden aprender sin estar programadas para hacerlo. En una

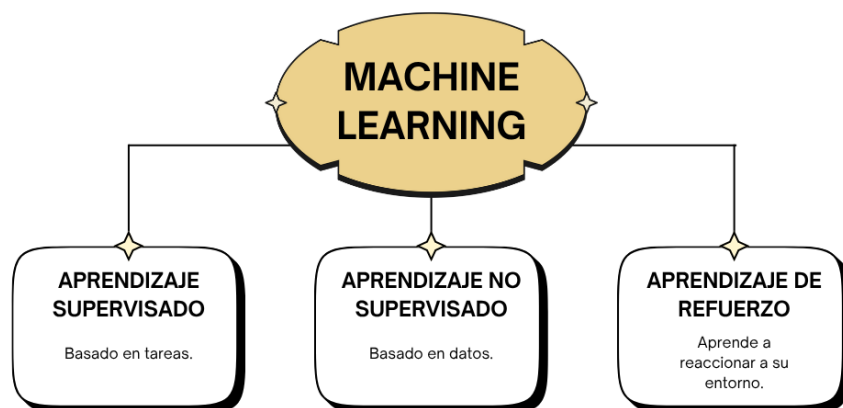
situación particular, las sugerencias o predicciones serían un resultado típico. (Rouhiainen, 2018)

Las primeras computadoras personales, que se hicieron accesibles a los consumidores a partir de la década de 1980, estaban programadas explícitamente para realizar ciertas tareas. Por el contrario, muchos de los dispositivos que verás en el futuro obtendrán experiencia y conocimientos de la forma en que se utilizan para poder brindar experiencias personalizadas al usuario gracias al aprendizaje automático. En la actualidad, la personalización de sitios de medios sociales como Facebook y los resultados del motor de búsqueda de Google son ejemplos de ello. (Rouhiainen, 2018)

Los algoritmos ayudan al aprendizaje automático a aprender patrones de datos. Los filtros de spam de correo electrónico, por ejemplo, utilizan este tipo de aprendizaje para identificar qué mensajes son spam y distinguirlos de los que no lo son. Este ejemplo ilustra cómo los algoritmos pueden usarse para aprender patrones y utilizar ese conocimiento para tomar decisiones. (Rouhiainen, 2018)

El aprendizaje automático supervisado, no supervisado y de refuerzo son los tres subconjuntos que se pueden usar, según la siguiente figura. (Rouhiainen, 2018)

Figura 4

Tipos de aprendizaje automático

Nota. Adaptado de Inteligencia Artificial [Fotografía], por Lasse Petteri, Rouhiainen, 2018, Editorial Planeta S.A.

3.2.11 Análisis de datos (Data Analysis)

El análisis de datos es un campo multidisciplinario que utiliza una variedad de técnicas de análisis, como matemáticas, estadística e informática, para extraer información de conjuntos de datos. El término "análisis de datos" abarca todo, desde el simple análisis de datos hasta la creación de teorías sobre cómo recopilarlos y almacenarlos. (Science, 2022)

a. ¿Cómo se utiliza el análisis de datos?

Los datos se utilizan todos los días, sin darse cuenta o no. Las tareas diarias, como medir los granos de café para preparar la taza de café de la mañana, consultar el informe meteorológico antes de decidir qué ponerse o utilizar un rastreador de actividad física, pueden analizar y utilizar datos. (Science, 2022)

b. Análisis de datos

Según (Science, 2022), nos menciona que “el análisis de datos descriptivo, diagnóstico, predictivo y prescriptivo son los cuatro tipos principales. Estos cuatro tipos de análisis de datos pueden ayudar a una organización a tomar decisiones basadas en datos cuando se utilizan juntos”. De un vistazo, cada uno de ellos nos dice lo siguiente:

- ✓ Los análisis descriptivos nos dicen lo que pasó.
- ✓ Los análisis de diagnóstico nos dicen por qué sucedió algo.
- ✓ El análisis predictivo nos ayuda a predecir lo que probablemente ocurrirá en el futuro.
- ✓ La analítica prescriptiva establece nuestras acciones.

Las personas que trabajan en análisis de datos generalmente explorarán cada una de estas cuatro áreas utilizando el proceso de análisis de datos, que incluye identificar la pregunta, recopilar datos sin procesarlos, limpiarlos, analizarlos e interpretar los resultados. (Science, 2022)

3.2.12 Redes neuronales (Neural Networks)

Una red neuronal es un programa o modelo de aprendizaje automático que toma decisiones de manera similar al cerebro humano, mediante el uso de métodos similares a cómo funcionan las neuronas biológicas trabajan juntas para identificar fenómenos, sopesar opciones y llegar a conclusiones. (Salas, 2004)

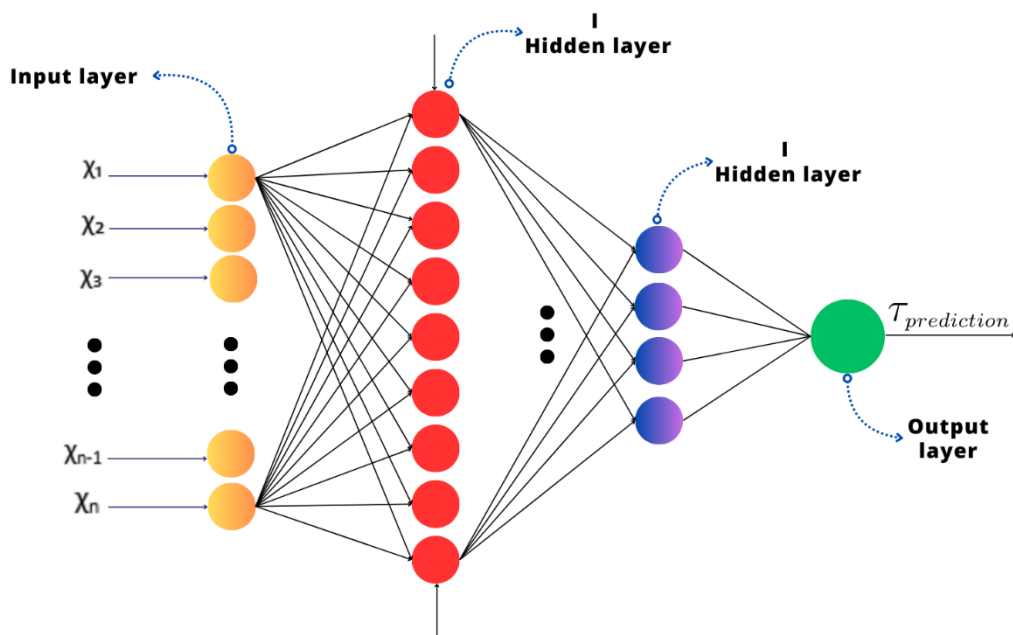
Cada red neuronal está compuesta por nodos o neuronas artificiales, que tienen una capa de entrada, una o más capas ocultas y una capa de salida. Cada nodo está conectado con otros y tiene un peso y un umbral asociados. El nodo se activa y envía datos a la

siguiente capa de la red si la salida de cualquier nodo está por encima del valor umbral establecido. De lo contrario, los datos no se transmiten a la capa posterior de la red. (Salas, 2004)

Los datos de entrenamiento son esenciales para que las redes neuronales aprendan y mejoren con el tiempo. Se convierten en herramientas poderosas en informática e inteligencia artificial una vez que se ajustan para lograr precisión, lo que nos permite clasificar y agrupar datos a alta velocidad. En comparación con la identificación manual realizada por expertos humanos, las tareas de reconocimiento de voz o imágenes pueden llevar minutos en lugar de horas. El algoritmo de búsqueda de Google es uno de los ejemplos de red neuronal más conocidos. (Salas, 2004)

Las redes neuronales se pueden referirse a veces como redes neuronales artificiales (ANN) o redes neuronales simuladas (CNN). Son fundamentales para los modelos de aprendizaje profundo y son un subconjunto del aprendizaje automático. (Salas, 2004)

Figura 5

Modelo de red neuronal

Nota. Adaptado de What is a Neural Network, [Fotografía], por IBM, International Business Machines, 2022.

a. Tipos de redes neuronales

Según (Salas, 2004), menciona que las redes neuronales se dividen en varios tipos y se utilizan para una variedad de propósitos. Aunque esta lista no incluye todos los tipos de redes neuronales, la siguiente sería una representación de los tipos más comunes que encontrará en sus casos de uso típicos:

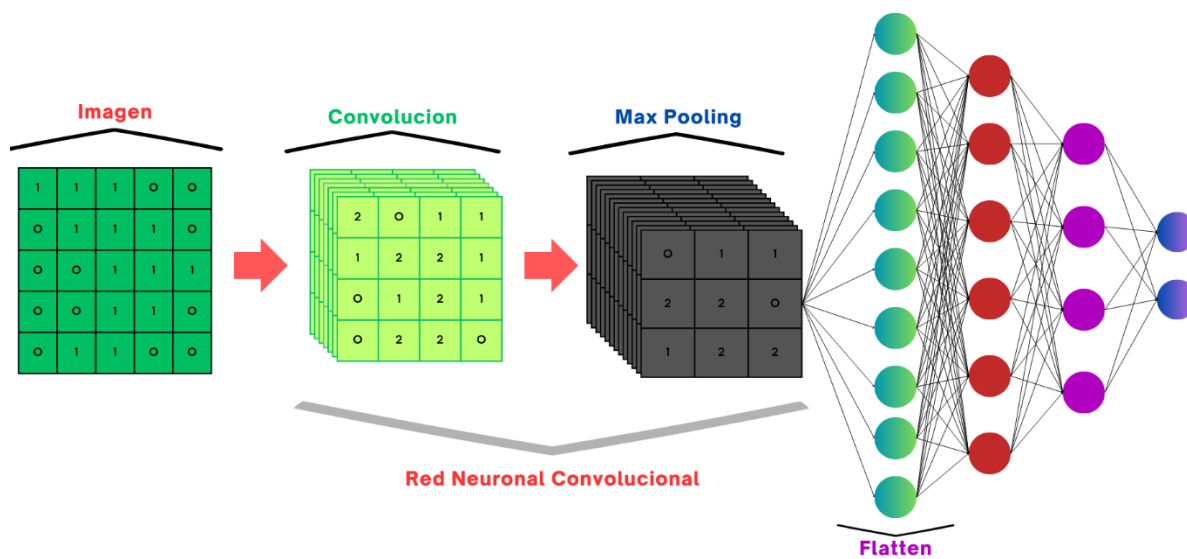
- ✓ En 1958, Frank Rosenblatt inventó el perceptrón, que es la red neuronal más antigua.

Nos indica (Salas, 2004) que:

Las redes neuronales de avance, también conocidas como perceptrones multicapa (MLP), son el tema principal de este artículo. Una capa de entrada, una capa o capas ocultas y una capa de salida son sus componentes. Aunque estas redes neuronales también se conocen como MLP, es importante señalar que, debido a que la mayoría de los problemas en el mundo real no son lineales, en realidad están compuestas de neuronas sigmoideas en lugar de perceptrones. Estos modelos generalmente se entrenan con datos y son la base de la visión por computadora, el procesamiento del lenguaje natural y otras redes neuronales.

- ✓ **Las redes neuronales convolucionales (CNN)** son similares a las redes feedforward, pero se utilizan principalmente para el reconocimiento de imágenes, el reconocimiento de patrones y/o la visión por computadora. Estas redes encuentran patrones dentro de una imagen utilizando la multiplicación de matrices. (Salas, 2004)

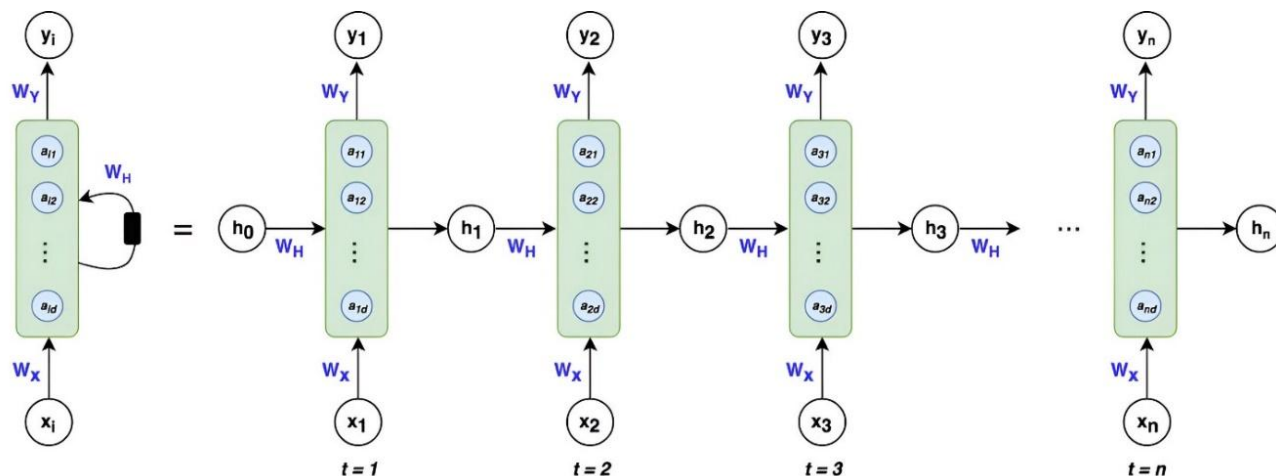
Figura 6

Estructura de redes neuronales convolucionales (CNN)

Nota. Adaptado de What is a Neural Network, [Fotografía], por IBM, International Business Machines, 2022.

- ✓ **Las redes neuronales recurrentes (RNN)** se identifican a través de sus bucles de comentarios. Se utilizan principalmente estos algoritmos de aprendizaje para hacer predicciones sobre resultados futuros, como predicciones del mercado de valores o pronósticos de ventas, con datos de series de tiempo. (Salas, 2004)

Figura 7

Estructura de redes neuronales recurrentes RNN

Nota. Adaptado de What is a Neural Network, [Fotografía], por IBM, International Business Machines, 2022.

3.2.13 Modelo de predicción (Predictive Modeling)

Los resultados se predicen utilizando estadísticas en el modelado predictivo. El modelado predictivo se puede aplicar a cualquier tipo de evento desconocido, pero la mayoría de las veces, el evento que uno quiere predecir es en el futuro. (Finlay, 2014)

En muchos casos, el modelo se elige basándose en la teoría de detección para intentar adivinar la probabilidad de un resultado dada una cantidad específica de datos de entrada; por ejemplo, para determinar la probabilidad de que un correo electrónico sea spam. (Finlay, 2014)

Los modelos pueden intentar determinar la probabilidad de que un conjunto de datos pertenezca a otro conjunto utilizando uno o más clasificadores. Por ejemplo, se podría usar

un modelo para determinar si un correo electrónico es spam o "ham" (no spam). (Finlay, 2014)

El modelado predictivo puede ser confundido o superpuesto en gran medida con el campo del aprendizaje automático, más conocido en los campos académicos o de investigación y desarrollo, dependiendo de los límites de las definiciones. El modelado predictivo con frecuencia se denomina análisis predictivo cuando se utiliza comercialmente. (Finlay, 2014)

El autor (Finlay, 2014) menciona que:

a. Aplicaciones

- ✓ Modelado de elevación
- ✓ Arqueología
- ✓ Gestión de relaciones con el cliente
- ✓ Seguro de automóvil
- ✓ Cuidado de la salud
- ✓ Comercio algorítmico
- ✓ Sistema de seguimiento de cliente potencial

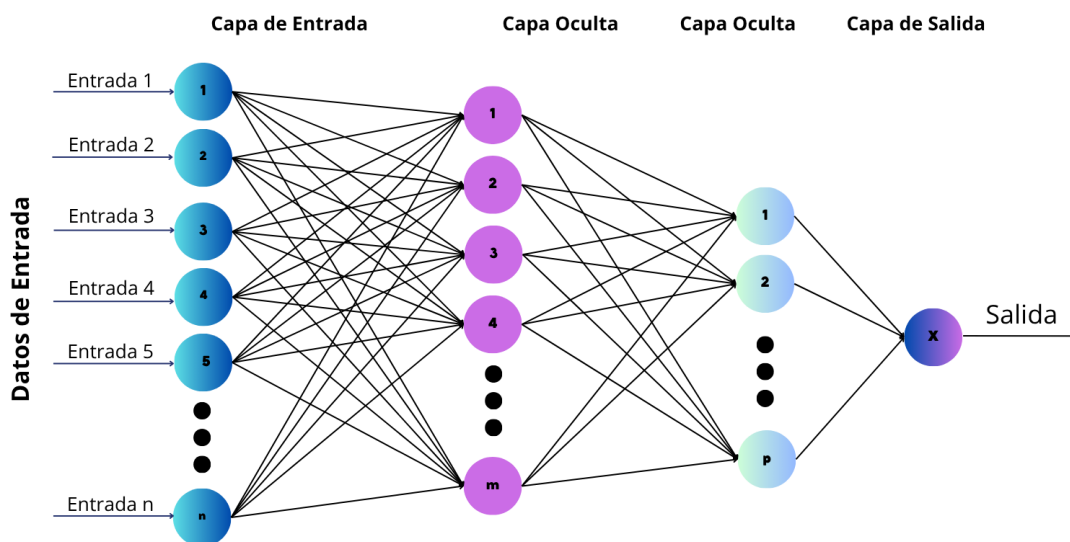
3.2.13 Aprendizaje profundo (*Deep Learning*)

El aprendizaje profundo es una de las aplicaciones más poderosas y de mayor crecimiento de la Inteligencia Artificial. Se trata de una rama del aprendizaje automático que se utiliza para resolver problemas muy complejos que generalmente implican grandes cantidades de datos. (Bengio y Hinton, 2015)

Las redes neuronales se organizan en capas para reconocer relaciones y patrones complejos en los datos, lo que permite el aprendizaje profundo. Su aplicación requiere una gran cantidad de datos y una gran capacidad de procesamiento. El reconocimiento de voz, el procesamiento del lenguaje natural, la visión artificial y la identificación de vehículos en los sistemas de asistencia al conductor son ejemplos de aplicaciones actuales. (Bengio y Hinton, 2015)

Las traducciones realizadas en Facebook son un buen ejemplo, ya que recientemente reveló que podía realizar alrededor de 4.500 millones de traducciones diarias gracias a un aprendizaje intensivo. Según (Bengio y Hinton, 2015), mencionan que “estos suelen ser pequeños fragmentos de texto, como las actualizaciones de estado que los usuarios publican en sus perfiles. Si no se lleva a cabo un aprendizaje profundo, sería extremadamente costoso y requeriría un equipo de personas muy grande para brindar el mismo servicio”.

Figura 8

Estructura de aprendizaje profundo

Nota. Adaptado de The Quartz guide to artificial intelligence: What is it, why is it important, and should we be afraid? [Fotografía], por Gershgorn, 2017.

3.2.14 Métricas de evaluación

3.2.15.1 Error Absoluto Medio [Mean Absolute Error (MAE)]. Según (Wang y Shanyou, 2023) menciona que “es una métrica muy utilizada en el análisis de regresión para examinar la eficacia de un modelo predictivo. Él cuantifica la disparidad absoluta promedio entre los valores observados y pronosticados dentro de un conjunto de datos”. La fórmula para MAE es la siguiente:

$$MAEc = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \text{-----> Ec. 1}$$

3.2.14.2 Error Medio Reducido [Mean Squared Error (MSE)]. Según (Wang y Shanyou, 2023), nos indica que “es una medida empleada frecuentemente en el análisis de regresión para evaluar la eficacia de un modelo predictivo”. Calcula el promedio de las discrepancias al cuadrado, o residuales, entre los valores observados y predichos dentro de un conjunto de datos. La fórmula para MSE es la siguiente:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{-----> Ec. 2}$$

3.2.14.3 Error Cuadrático Medio [Root Mean Squared Error (RMSE)]. Según (Wang y Shanyou, 2023), menciona que “es una medición común en el análisis de regresión que se utiliza para evaluar la eficacia de los modelos predictivos. Surge del error cuadrático medio (MSE) y ofrece una indicación de la magnitud típica de los errores o residuos entre los valores observados y predichos, considerando tanto su dirección como su magnitud”. RMSE se calcula tomando la raíz cuadrada de MSE. La fórmula para RMSE se deriva de lo siguiente:

$$RMSE = \sqrt{MSE} \quad \text{-----> Ec. 3}$$

3.2.14.4 Coeficiente de Determinación R^2 [Coefficient of determination R^2 (R^2 -square)]. Según (Wang y Shanyou, 2023), el análisis de regresión, el coeficiente de determinación, también conocido como R-cuadrado, es una métrica crucial que se utiliza para evaluar la bondad del ajuste de un modelo de regresión. Representa la proporción de la varianza en la variable dependiente que las variables independientes han predeterminado. El valor varía de 0 a 1, donde:

- a) $R^2 = 0$ indica que el modelo no proporciona ninguna explicación sobre la variabilidad de la variable dependiente.
- b) $R = 1$ indica que el modelo proporciona una explicación completa de la variabilidad de la variable dependiente.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{-----> Ec. 4}$$

3.2.14.5 Error de Porcentaje Absoluto Medio [Mean Absolute Percentage Error (MAPE)]. Según (Wang y Shanyou, 2023), es “una métrica utilizada con frecuencia para medir la precisión de los modelos de pronóstico, particularmente en el análisis de series de tiempo”.

Se utiliza un porcentaje para calcular la diferencia porcentual absoluta promedio entre los valores reales y pronosticados. La fórmula de MAPE es:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100\% \quad \text{-----> Ec. 5}$$

3.2.14.6 Error Porcentual Absoluto Medio Simétrico [Symmetric Mean Absolute Percentage Error (SMAPE)]. Según (Wang y Shanyou, 2023), es “una métrica que se utiliza para evaluar la precisión de los pronósticos, particularmente en tareas de pronóstico y análisis de series de tiempo”. Mide la diferencia porcentual entre los valores previstos y reales, teniendo en cuenta la magnitud de los valores pronosticados. A menudo se prefiere SMAPE cuando se trata de datos que contienen valores atípicos o cuando la escala de los datos varía ampliamente.

La fórmula de SMAPE es:

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|f_t - A_t|}{(|f_t| + |A_t|)/2} \quad \text{-----> Ec. 6}$$

Donde:

F_t = valor previsto.

A_t = valor real.

n = número total de observaciones.

SMAPE expresa la precisión del pronóstico como un porcentaje, donde valores más bajos indican una mayor precisión. Es de naturaleza simétrica, lo que significa que no favorece ni las subestimaciones ni las sobreestimaciones. Esta característica hace que SMAPE sea útil para comparar la precisión de varios modelos de pronóstico, especialmente cuando los valores pronosticados son de escalas diferentes.

3.2.15 Marco legal

a. Constitución Política del Perú

Inciso N°22 del Artículo 2, Capítulo I, Título I. La paz, la tranquilidad, el tiempo libre y el descanso, y un entorno equilibrado y adecuado para el desarrollo de su vida.

b. Ley N°28611, Ley General del Ambiente

Inciso N°1 del Artículo 31, Capítulo III, Título I. Estándar de Calidad Ambiental (ECA). Es la medida que determina la concentración o el grado de elementos, sustancias o parámetros físicos, químicos y biológicos presentes en el aire, agua o suelo en su condición de cuerpo receptor, que no representa un riesgo significativo para la salud de las personas ni para el ambiente. La concentración o el grado se pueden representar en máximos, mínimos o rangos según el parámetro en cuestión. (Minan, 2008)

c. D.S. N°004-2017-MINAM

Aprueban Estándares de Calidad Ambiental (ECA) para Agua y establecen Disposiciones Complementarias. Para preservar los recursos hídricos y fomentar el desarrollo sostenible, se establecen límites para no excederlos. Los Estándares Nacionales de Calidad Ambiental se clasifican en cuatro categorías, y las 37 normas se enumeran en cada una de ellas. (Minan, 2008)

El propósito de esta norma es reunir las disposiciones emitidas por el Decreto Supremo N°002-2008-MINAM, el Decreto Supremo N°023-2009-MINAM y el Decreto Supremo N°015-2015-MINAM, que establecen los Estándares de Calidad Ambiental (ECA) para Agua. Todas estas disposiciones están sujetas a lo establecido en el presente Decreto Supremo y el Anexo que lo acompaña. Esta recopilación de leyes altera y elimina ciertos valores, parámetros, categorías y subcategorías de los ECA, mientras que se mantienen otros que fueron aprobados por los decretos supremos mencionados anteriormente. (Minan, 2008)

d. Categoría 1. Poblacional y recreacional

Los estándares de calidad ambiental (ECA); la Organización Mundial de la Salud (OMS), la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) evalúan la concentración de elementos minerales, materias orgánicas e inorgánicas, entre otros, presentes en lagunas, lagos y ríos en todo el Perú, en los que sus aguas se utilizan para diferentes propósitos. (Diario el Peruano, 2017)

Estudios actuales realizados por la OMS, la EPA y la FAO donde cumple con criterios. a través de lo cual Desde que el Ministerio del Medio Ambiente (MINAM)

modificó los estándares de calidad del agua, los minerales como el cromo y el arsénico ya se han considerado por los ECA anteriores para aprobar este tipo de agua en 2008. Los valores de los parámetros para ciertos usos aumentaron como resultado de cambios en las normas internacionales sobre calidad ambiental del agua. El MINAM comparte el proceso y el conjunto de esta regla. (Minan, 2008)

El estándar de calidad ambiental se define como: una medida que establece el nivel de concentración o el grado de elementos, sustancias en el agua, el aire y el suelo en función de valores de parámetros físicos, químicos y biológicos, considerando que no representa un peligro para la salud de la población y el ambiente. El estándar de calidad ambiental es una unidad de medida que mide la calidad al uso que se le da al cuerpo de agua, donde influyen dos factores: la carga contaminante y su valor natural. (Diario el Peruano, 2017)

En Perú, a partir de la Ley de Aguas (Decreto Ley N°17752 de 1969) y posteriormente la Ley de Recursos Hídricos (Ley N°29338 de 2009), se estableció que los ECA de agua deben establecerse en función de las categorías de agua utilizadas para el uso del cuerpo natural, como se indica a continuación:

Tabla 2

Categorías en relación al uso para evaluar los parámetros según ECA Agua

| Categoría | Descripción | Subcategoría | Descripción |
|------------------|-------------------------------------|---------------------|---|
| Categoría 1-A | Aguas superficiales destinadas a la | A1 | Agua que puede ser potabilizada con desinfección. |

| Categoría | Descripción | Subcategoría | Descripción |
|--|---|---------------------|--|
| Categoría 1-B | producción de agua potable. | A2 | Agua que puede ser potabilizada con tratamiento convencional. |
| | | A3 | Agua que puede ser potabilizada con tratamiento avanzado. |
| | | B1 | Contacto primario. |
| Categoría 1-B | Aguas superficiales destinadas a recreación | B2 | Contacto secundario. |
| | | C1 | Extracción y cultivo de moluscos bivalvos. |
| Categoría 2: Actividades de extracción y cultivo marino costeras y continentales. | Agua de mar | C2 | Extracción y cultivo de otras especies hidrobiológicas. |
| | | C3 | Otras actividades. |
| | | C4 | Extracción y cultivo de especies hidrobiológicas en lagos o lagunas. |
| Categoría 3: Riesgo de vegetales y | Parámetros para riesgo de vegetales. | D1 | Riesgo de cultivos de tallo alto y bajo. |
| | Parámetros para bebida de animales. | D2 | Bebida de animales. |

| Categoría | Descripción | Subcategoría | Descripción |
|---------------------|------------------------------------|-----------------------------|-------------------------|
| bebida de animales. | | | |
| Categoría 4: | Conservación del Ambiente Acuático | E1 | Lagunas y lagos. |
| | | E2: Ríos | Ríos de costa y sierra. |
| | | | Ríos de selva. |
| | | E3: | |
| | | Ecosistemas marino costeras | Estuarios. Marinos. |

Nota. Adaptado de Estándar de Calidad Ambiental AGUA [Tabla], por Ministerio del Ambiente, 2017.

Asimismo, es importante destacar que los ECA de Agua del Perú se han establecido considerando referentes internacionales.

a. Sub Categoría A: Aguas superficiales destinadas a la producción de agua potable

Se refiere a las aguas destinadas para el consumo humano después de haber sido tratadas. (Diario el Peruano, 2017)

b. Aguas que pueden ser potabilizadas con desinfección

Se refiere a las aguas que, de acuerdo con sus características de calidad, cumplen con los requisitos para ser destinadas al suministro de agua para consumo humano mediante desinfección simple, de acuerdo con las regulaciones vigentes. (Diario el Peruano, 2017)

Tabla 3

Categoría 1. Poblacional y recreacional, Sub Categoría A1 – ECA

| Parámetros | Unidad de Medida | A1 |
|--|--------------------|---|
| | | Aguas que pueden ser potabilizadas con desinfección |
| FISICOS - QUIMICOS | | |
| Aceites y grasas | mg/L | 0.5 |
| Cloruros | mg/L | 0.07 |
| Conductividad | (\square S/cm) | 1 500 |
| Demanda Bioquímica de Oxígeno (DBO) | mg/L | 3 |
| Dureza | mg/L | 500 |
| Demanda Química de Oxígeno (DQO) | mg/L | 10 |
| Nitratos | mg/L | 50 |
| Nitritos | mg/L | 3 |
| Potencial de Hidrógeno | Unidad de Ph | 6,5 - 8,5 |
| Sólidos Disueltos Totales | mg/L | 1 000 |
| Sulfatos | mg/L | 250 |
| Temperatura | °C | Δ 3 |
| Turbiedad | UNT | 5 |
| MICROBIOLÓGICOS Y PARASITOLÓGICOS | | |
| Coliformes Totales | NMP/100ml | 50 |
| Coliformes Termotolerantes | NMP/100ml | 20 |
| <i>Escherichia coli</i> | NMP/100ml | 0 |
| <i>Vibrio Cholerae</i> | Presencial / 100ml | Ausencia |

Nota. Adaptado de Estándar de Calidad – ECA Agua [Tabla], por Ministerio del Ambiente, 2017.

3.3 Definición de Términos

- ✓ **Agente.** Cualquier cosa física, química o biológica que pueda tener un efecto negativo.
- ✓ **Agente patológico.** Perteneciente o relacionado con la enfermedad, que indica o implica enfermedad. Son agentes infecciosos que pueden causar enfermedades a su huésped.
- ✓ **Ambiente acuático.** Lugares en cuerpos de agua. ríos, arroyos, lagos, lagunas, esteros y canales, por ejemplo.
- ✓ **Análisis determinísticos.** Análisis en el que se asume que todos los parámetros ambientales y poblacionales son constantes y se describen como tal.
- ✓ **Áreas de influencia.** El área cercana a la instalación donde hay indicios o pruebas de contaminación.
- ✓ **Comunidad.** Un conjunto de poblaciones de varias especies que viven juntas y interactúan.
- ✓ **Contaminación.** Distribución de una sustancia química o una mezcla de sustancias en un lugar no deseable (aire, agua, suelo), donde puede tener efectos perjudiciales para el medio ambiente o la salud.
- ✓ **Contaminante.** Cualquier sustancia química que no es natural del suelo o cuya concentración supera el nivel de fondo puede ser perjudicial para la salud de las personas o el medio ambiente.
- ✓ **Ecosistema.** Un conjunto dinámico de comunidades vegetales, animales y de microorganismos junto con su entorno no vivo que interactúan como una unidad funcional.

- ✓ **Estándar de Calidad Ambiental (ECA).** Es la medida que determina la concentración o el grado de elementos, sustancias o parámetros físicos, químicos y biológicos presentes en el aire, agua o suelo en su condición de cuerpo receptor, que no representa un riesgo significativo para la salud de las personas ni para el ambiente. La concentración o el grado se pueden representar en máximos, mínimos o rangos según el parámetro en cuestión.
- ✓ **Geo-referenciación.** Es el proceso técnico-científico de establecer la ubicación espacial de un objeto en un sistema de datum y coordenadas específicas.
- ✓ **GPS.** El sistema de posicionamiento global funciona mediante una red de satélites en órbita sobre el planeta y permite determinar la posición geográfica de un objeto, persona o nave en cualquier parte del mundo.
- ✓ **Inteligencia Artificial.** Es el conjunto de algoritmos desarrollados con el objetivo de desarrollar programas informáticos que tengan las mismas habilidades que los humanos.
- ✓ **Microorganismos patológicos.** Son bacterias que pueden provocar diferentes patologías a personas o animales que los ingieran.
- ✓ **Peligro.** Cuando las personas, sistemas o poblaciones están expuestos a un (o varios) agente(s) de estrés, tienen la capacidad de producir efecto(s) negativo(s).
- ✓ **Recursos renovables.** Son los recursos naturales que no han sido alterados por el ser humano.

IV. Metodología

4.1 Tipo y nivel de investigación

4.1.1 Tipo de investigación

El estudio de investigación presenta un enfoque cuantitativo, el tipo de investigación es aplicada, porque su finalidad principal es resolver un problema práctico y real en un contexto específico. El estudio busca abordar una problemática concreta en la cuenca del río Chumbao por la presencia de agentes patológicos. Al utilizar Inteligencia Artificial para la predicción de agentes patológicos, el conocimiento generado permite a las autoridades de Andahuaylas tomar decisiones preventivas y sanitarias directas sobre el recurso hídrico. La investigación toma teorías ya probadas como parámetros de calidad del agua y modelos de IA y se aplica a un escenario geográfico y temporal delimitado para generar una herramienta funcional.

4.1.2 Nivel de investigación

Según su naturaleza o profundidad, el nivel de una investigación se refiere al grado de conocimiento que posee el investigador en relación con el problema, hecho o fenómeno a estudiar. De igual modo cada nivel de investigación emplea estrategias adecuadas para llevar a cabo el desarrollo de la investigación (Condori, 2020).

El nivel de investigación para el presente estudio es correlacional – predictivo porque, el estudio busca primero cuantificar la relación entre los parámetros de calidad del agua con los parámetros físico-químicos y la presencia de agentes patológicos con los parámetros microbiológicos. Al incorporar Inteligencia Artificial, el objetivo trasciende, porque existe una correlación sólida, la cual proporciona una base para la predicción. Los modelos de IA aprenden el comportamiento de las variables para anticipar la aparición de patógenos sin necesidad de realizar muestreos físicos y costosos en cada punto.

En resumen, es correlacional porque asocia variables y es predictivo porque utiliza esa asociación para estimar agentes patológicos.

4.2 Diseño de investigación

Los diseños de investigación son esquemas o enfoques metodológicos que se emplean para dirigir la recopilación y el análisis de datos en un estudio. A continuación, se muestra los diseños de investigación que se seleccionó para nuestra investigación. (Hernández y Fernandèz, 2014)

La investigación emplea un diseño no experimental de corte longitudinal, fundamentado en el registro histórico de calidad del agua en la cuenca hidrográfica del río Chumbao. Los datos proceden de diversas instituciones locales y nacionales, en distintos periodos temporales permitiendo evaluar los parámetros físico-químicos para la predicción de parámetros microbiológicos.

4.3 Ámbito temporal y espacial

4.3.1 Población y muestra

Población. La población para el presente estudio debido a la necesidad del uso de datos a procesarse en los modelos de inteligencia artificial se ha considerado a los registros de monitoreo de calidad del agua existentes para la cuenca hidrográfica del río Chumbao; mismos que fueron generados por las distintas instituciones locales y nacionales vinculadas a la gestión de recurso hídrico. Dentro de las cuales tenemos: DIRESA, ANA y la EPS EMSAP CHANKA.

Figura 9

Población de la cuenca del río Chumbao



Nota. Imagen de Google Earth, 2024.

Muestra. El tipo de muestra es no probabilística – convencional, ya que se seleccionaron registros de monitoreos de calidad del agua que contenían información con respecto a parámetros microbiológicos, parasitológicos y parámetros físico-químicos, haciendo un total de 126 registros para el procesamiento de datos en la predicción de los modelos de inteligencia artificial. Como se muestra en la Tabla 4.

Tabla 4

Puntos de muestreo

| Muestra | Control | Puntos de Muestreo | Coordenadas UTM WGS 84 – ZONA 18S | |
|-----------|---------|--------------------|-----------------------------------|----------------|
| | | | Este | Norte |
| Muestra 1 | | P1 | 682457.00 m E | 8483792.00 m S |
| | | P2 | 676659.00 m E | 8489942.00 m S |

| Muestra | Control | Puntos de Muestreo | Coordenadas UTM WGS 84 – ZONA 18S | |
|------------------|------------------|--------------------|-----------------------------------|----------------|
| | | | Este | Norte |
| | | P3 | 673857.00 m E | 8489386.00 m S |
| | | P4 | 669719.00 m E | 8489597.00 m S |
| | | P5 | 667376.00 m E | 8491705.00 m S |
| | | P1 | 682141.00 m E | 8484763.00 m S |
| | Control 1 | P2 | 672369.00 m E | 8489244.00 m S |
| | | P3 | 669757.00 m E | 8489581.00 m S |
| | | P4 | 667354.00 m E | 8491765.00 m S |
| | | P1 | 682141.00 m E | 8484763.00 m S |
| Muestra 2 | Control 2 | P2 | 672369.00 m E | 8489244.00 m S |
| | | P3 | 669757.00 m E | 8489581.00 m S |
| | | P4 | 667354.00 m E | 8491765.00 m S |
| | | P1 | 682141.00 m E | 8484763.00 m S |
| | Control 3 | P2 | 672369.00 m E | 8489244.00 m S |
| | | P3 | 669757.00 m E | 8489581.00 m S |
| | | P4 | 667354.00 m E | 8491765.00 m S |
| | | P2 | 679846.00 m E | 8486703.00 m S |
| | | P5 | 676686.00 m E | 8490248.00 m S |
| | | P6 | 673710.00 m E | 8489650.00 m S |
| | Control 1 | P7 | 673550.00 m E | 8489661.00 m S |
| | | P8 | 670930.00 m E | 8489630.00 m S |
| Muestra 3 | | P9 | 668300.00 m E | 8490680.00 m S |
| | | P10 | 667468.00 m E | 8492261.00 m S |
| | | P2 | 679846.00 m E | 8486703.00 m S |
| | Control 2 | P5 | 676686.00 m E | 8490248.00 m S |
| | | P6 | 673710.00 m E | 8489650.00 m S |
| | | P7 | 673550.00 m E | 8489661.00 m S |

| Muestra | Control | Puntos de Muestreo | Coordenadas UTM WGS 84 – ZONA 18S | |
|------------------|------------------|--------------------|-----------------------------------|----------------|
| | | | Este | Norte |
| | | P8 | 670930.00 m E | 8489630.00 m S |
| | | P9 | 668300.00 m E | 8490680.00 m S |
| | | P10 | 667468.00 m E | 8492261.00 m S |
| | | P2 | 679846.00 m E | 8486703.00 m S |
| | | P5 | 676686.00 m E | 8490248.00 m S |
| | | P6 | 673710.00 m E | 8489650.00 m S |
| | Control 3 | P7 | 673550.00 m E | 8489661.00 m S |
| | | P8 | 670930.00 m E | 8489630.00 m S |
| | | P9 | 668300.00 m E | 8490680.00 m S |
| | | P10 | 667468.00 m E | 8492261.00 m S |
| | | P2 | 679656.00 m E | 8486343.00 m S |
| | | P3 | 677545.00 m E | 8489661.00 m S |
| | | P4 | 675515.00 m E | 8489374.00 m S |
| Muestra 4 | $\bar{a} \pm S$ | P5 | 673275.00 m E | 8489272.00 m S |
| | | P6 | 669728.00 m E | 8489590.00 m S |
| | | P7 | 667344.00 m E | 8491769.00 m S |
| | | P8 | 667672.00 m E | 8497010.00 m S |
| Muestra 5 | | P1 | 674861.00 m E | 8489569.00 m S |
| Muestra 6 | | P1 | 671141.00 m E | 8489142.00 m S |
| | | P4 | 682451.70 m E | 8483785.89 m S |
| | | P5 | 677536.68 m E | 8489658.96 m S |
| | | P6 | 675509.18 m E | 8489377.57 m S |
| Muestra 7 | Control 1 | P7 | 673374.71 m E | 8489271.90 m S |
| | | P8 | 669704.43 m E | 8489603.22 m S |
| | | P9 | 667382.79 m E | 8491710.93 m S |
| | | P10 | 667663.02 m E | 8497016.42 m S |

| Muestra | Control | Puntos de Muestreo | Coordenadas UTM WGS 84 – ZONA 18S | |
|------------------|------------------|--------------------|-----------------------------------|----------------|
| | | | Este | Norte |
| | | P4 | 682451.70 m E | 8483785.89 m S |
| | | P5 | 677536.68 m E | 8489658.96 m S |
| | | P6 | 675509.18 m E | 8489377.57 m S |
| | Control 2 | P7 | 673374.71 m E | 8489271.90 m S |
| | | P8 | 669704.43 m E | 8489603.22 m S |
| | | P9 | 667382.79 m E | 8491710.93 m S |
| | | P10 | 667663.02 m E | 8497016.42 m S |
| | | P4 | 682451.70 m E | 8483785.89 m S |
| | | P5 | 677536.68 m E | 8489658.96 m S |
| | | P6 | 675509.18 m E | 8489377.57 m S |
| | Control 3 | P7 | 673374.71 m E | 8489271.90 m S |
| | | P8 | 669704.43 m E | 8489603.22 m S |
| | | P9 | 667382.79 m E | 8491710.93 m S |
| | | P10 | 667663.02 m E | 8497016.42 m S |
| Muestra 8 | | P1 | 675564.00 m E | 8489373.00 m S |
| | | P1 | 682156.00 m E | 8484810.00 m S |
| | | P2 | 682156.00 m E | 8484810.00 m S |
| | | P3 | 682156.00 m E | 8484810.00 m S |
| | | P4 | 682156.00 m E | 8484810.00 m S |
| | | P5 | 682156.00 m E | 8484810.00 m S |
| | | P6 | 682156.00 m E | 8484810.00 m S |
| Muestra 9 | Control 1 | P7 | 682156.00 m E | 8484810.00 m S |
| | | P8 | 682156.00 m E | 8484810.00 m S |
| | | P9 | 682156.00 m E | 8484810.00 m S |
| | | P10 | 682156.00 m E | 8484810.00 m S |
| | | P11 | 682156.00 m E | 8484810.00 m S |
| | | P12 | 682156.00 m E | 8484810.00 m S |

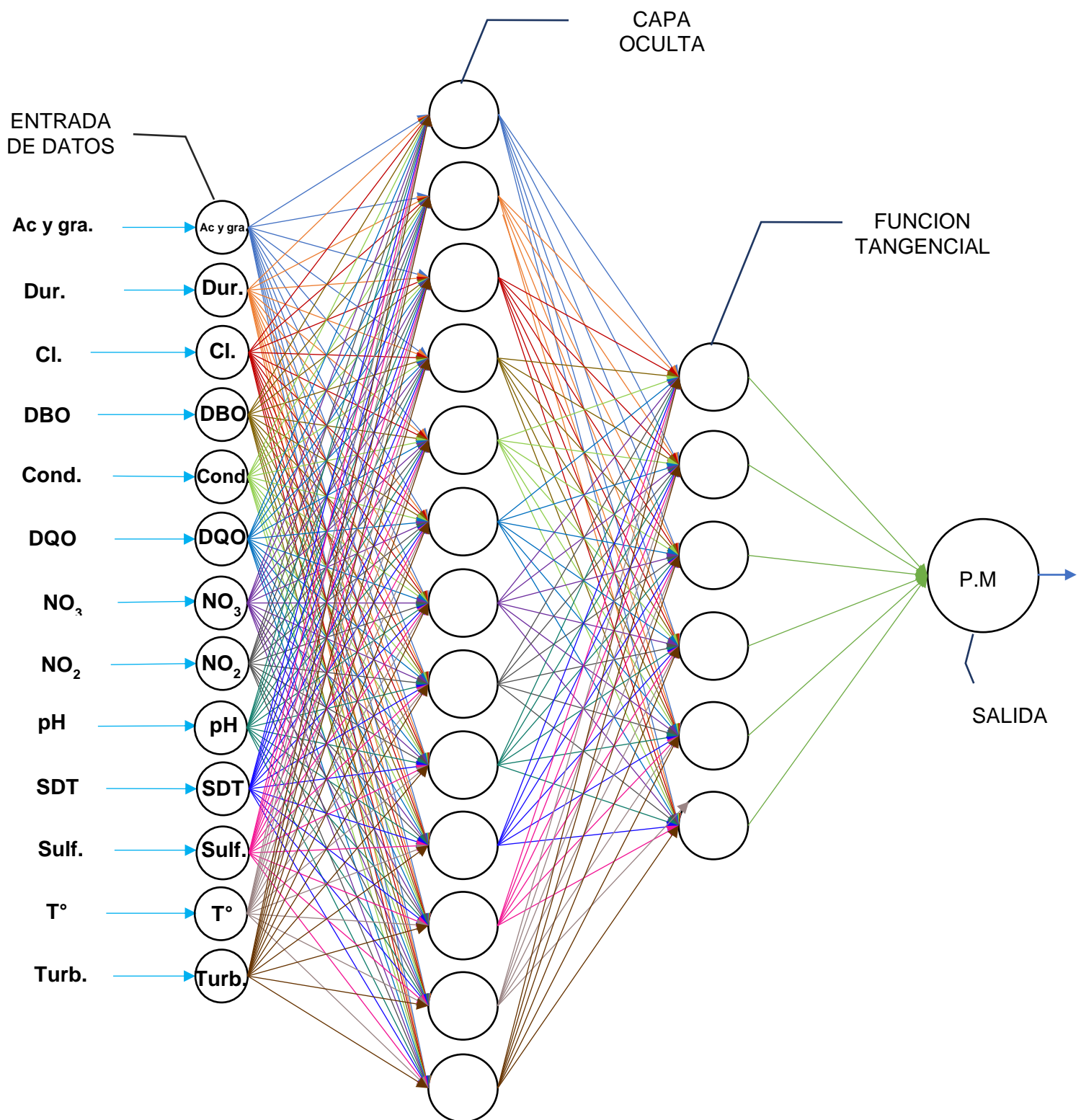
| Muestra | Control | Puntos de Muestreo | Coordenadas UTM WGS 84 – ZONA 18S | |
|-------------------|------------------|--------------------|-----------------------------------|----------------|
| | | | Este | Norte |
| | Control 2 | P1 | 682138.00 m E | 8484774.00 m S |
| | | P2 | 682138.00 m E | 8484774.00 m S |
| | | P3 | 682138.00 m E | 8484774.00 m S |
| | | P4 | 682138.00 m E | 8484774.00 m S |
| | | P5 | 682138.00 m E | 8484774.00 m S |
| | | P6 | 682138.00 m E | 8484774.00 m S |
| | | P7 | 682138.00 m E | 8484774.00 m S |
| | | P8 | 682138.00 m E | 8484774.00 m S |
| | | P9 | 682138.00 m E | 8484774.00 m S |
| | | P10 | 682138.00 m E | 8484774.00 m S |
| | | P11 | 682138.00 m E | 8484774.00 m S |
| | | P12 | 682138.00 m E | 8484774.00 m S |
| | Control 3 | P1 | 682126.00 m E | 8484806.00 m S |
| | | P2 | 682126.00 m E | 8484806.00 m S |
| | | P3 | 682126.00 m E | 8484806.00 m S |
| | | P4 | 682126.00 m E | 8484806.00 m S |
| | | P5 | 682126.00 m E | 8484806.00 m S |
| | | P6 | 682126.00 m E | 8484806.00 m S |
| | | P7 | 682126.00 m E | 8484806.00 m S |
| | | P8 | 682126.00 m E | 8484806.00 m S |
| | | P9 | 682126.00 m E | 8484806.00 m S |
| | | P10 | 682126.00 m E | 8484806.00 m S |
| | | P11 | 682126.00 m E | 8484806.00 m S |
| | | P12 | 682126.00 m E | 8484806.00 m S |
| Muestra 10 | | P1 | 672571.00 m E | 8489093.00 m S |
| | | P2 | 673418.00 m E | 8489886.00 m S |
| | | P3 | 672635.00 m E | 8489710.00 m S |

| Muestra | Control | Puntos de Muestreo | Coordenadas UTM WGS 84 – ZONA 18S | |
|-------------------|---------|--------------------|-----------------------------------|----------------|
| | | | Este | Norte |
| Muestra 11 | | P1 | 691361.00 m E | 8475680.00 m S |
| | | P3 | 692174.00 m E | 8472786.00 m S |
| | | P4 | 690128.00 m E | 8479187.00 m S |
| | | P5 | 691327.00 m E | 8476615.00 m S |
| Muestra 12 | | P1 | 677021.00 m E | 8488682.00 m S |
| | | P2 | 676989.00 m E | 8488922.00 m S |
| | | P3 | 677022.00 m E | 8489130.00 m S |
| Muestra 13 | | P1 | 677127.00 m E | 8489872.00 m S |
| | | P2 | 675625.00 m E | 8489376.00 m S |
| | | P3 | 668990.00 m E | 8489734.00 m S |
| Muestra 14 | | P1 | 669446.00 m E | 8489690.00 m S |
| Muestra 15 | | P1 | 669585.00 m E | 8489662.00 m S |
| Muestra 16 | | P1 | 671513.00 m E | 8490678.00 m S |
| | | P2 | 669632.00 m E | 8489984.00 m S |
| | | P3 | 669669.00 m E | 8489889.00 m S |
| | | P4 | 669762.00 m E | 8489935.00 m S |
| | | P5 | 669565.0 m E | 8489924.00 m S |
| | | P6 | 669585.00 m E | 8489662.00 m S |
| | | | Total, de puntos: 126 | |

Nota. Elaboración propia

Muestreo. El tipo de muestreo es por conveniencia, ya que, el muestreo fue deliberado, crítico o por juicio la cual consiste en seleccionar una muestra basada en nuestros conocimientos sobre el tema.

Figura 10

Arquitectura Propuesta de una Red Neuronal

Nota. La presente gráfica representa la arquitectura de una red neuronal que tendrá como datos de entrada los parámetros físico químicos, tendrá un número finito de capas ocultas, y como datos de salida los parámetros microbiológicos.

Donde:**Datos de entrada:**

- ✓ **Ac y Gra.** = Aceites y grasas
- ✓ **Cl.** = Cloruros
- ✓ **Cond.** = Conductividad
- ✓ **DBO** = Demanda Bioquímica de Oxígeno
- ✓ **Dur.** = Dureza
- ✓ **DQO** = Demanda Química de Oxígeno
- ✓ **NO3** = Nitratos
- ✓ **NO2** = Nitritos
- ✓ **pH** = Potencial de Hidrógeno
- ✓ **SDT** = Sólidos Disueltos Totales
- ✓ **Sulf.** = Sulfatos
- ✓ **T°** = Temperatura
- ✓ **Turb.** = Turbiedad

Datos de salida:

- ✓ **P.M.** = Parámetros microbiológicos

4.4 Instrumentos

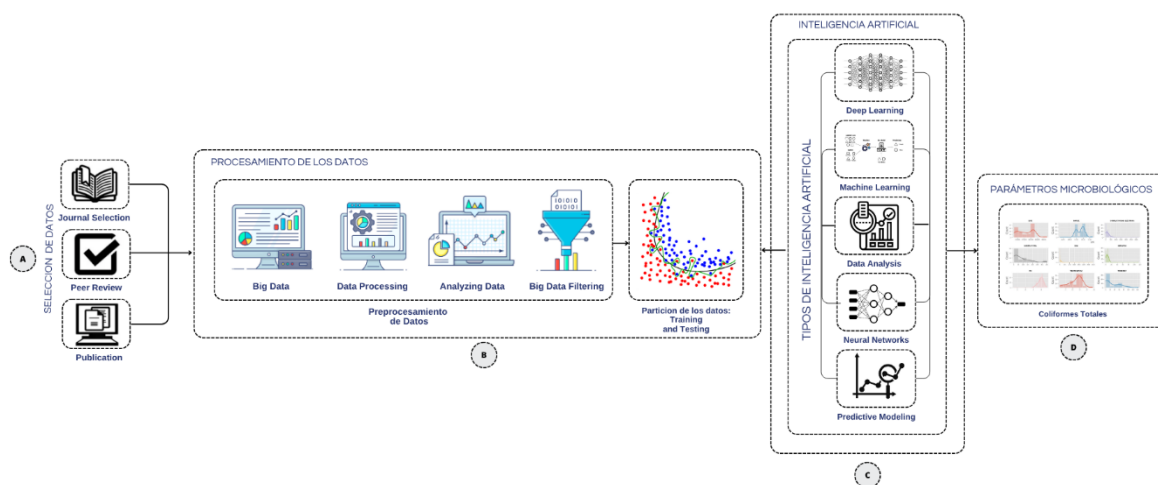
- ✓ Registro de datos de monitoreos de calidad del agua.
- ✓ Computadora.
- ✓ Modelos de inteligencia artificial.

4.5 Procedimientos

4.5.1 Diagrama de flujo del proyecto de investigación

Figura 11

Diagrama de flujo del proyecto de predicción de Coliformes Totales



Nota. Arquitectura del proyecto de investigación aplicando aprendizaje supervisado para la predicción de coliformes totales. Elaboración propia.

A. Selección de Datos

Los datos para la investigación se seleccionaron de diferentes artículos que realizaron estudios de campo en la cuenca hidrográfica del río Chumbao, de los cual se hizo un cuadro de los parámetros en común, para entrenar nuestros modelos de Red.

En la Tabla 4, se muestra los datos con los parámetros a utilizar en el modelo de red de Machine Learning, se tienen 10 características en común de los datos como son: la ubicación Este y norte, conductividad, dureza total, DBO, PH, Temperatura, Turbidez y Coliformes

totales, de los cuales los 9 parámetros iniciales serán nuestros datos de entrada y como salida tendremos el parámetro de coliformes totales que será nuestro predictivo.

Los datos que analizamos algunos no se encontraban con la misma cantidad, entonces se tuvo que rellenar utilizando interpolación o promedios, además estos datos también su valor números se difieren mucho en su medida entonces lo que se realizó fue escalar estos datos para que estén más homogéneos uno del otro.

Tabla 5

Datos químicos y biológicos

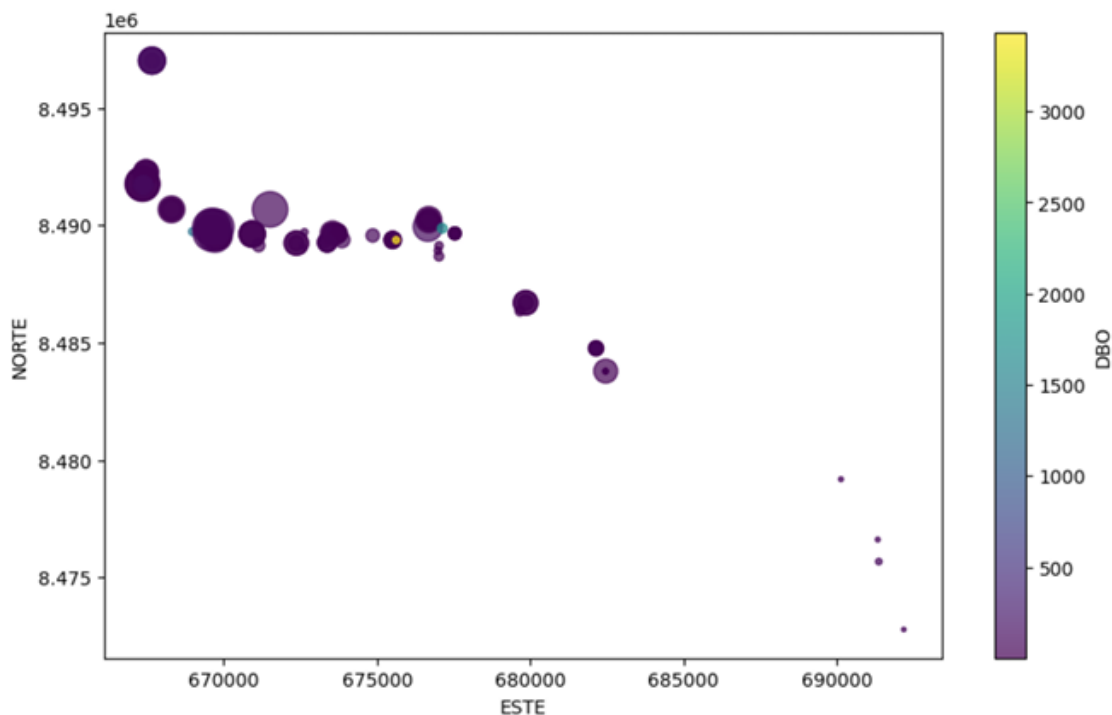
| Nº | Este | Norte | Cond. Eléctrica | Dureza Total | DBO | Nitratos | pH | T° | Turbidez | Coliformes Totales |
|------------|----------|-----------|-----------------|--------------|--------|----------|------|-----|----------|--------------------|
| 0 | 682457.0 | 8483792.0 | 161.0 | 159.0 | 5.40 | 4.90 | 9.30 | 4.0 | 90.0 | 300.0 |
| 1 | 676659.0 | 8489942.0 | 241.0 | 241.0 | 117.20 | 1.50 | 1.20 | 1.7 | 77.0 | 110000.0 |
| 2 | 673857.0 | 8489386.0 | 355.0 | 71.0 | 65.25 | 9.60 | 9.40 | 8.7 | 15.0 | 110000.0 |
| 3 | 669719.0 | 8489597.0 | 126.0 | 339.0 | 106.48 | 1.60 | 8.00 | 8.4 | 89.0 | 110000.0 |
| 4 | 667376.0 | 8491705.0 | 247.0 | 117.0 | 39.00 | 2.90 | 6.10 | 7.3 | 50.0 | 110000.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 121 | 669632.0 | 8489984.0 | 646.0 | 350.0 | 47.00 | 0.07 | 6.87 | 3.7 | 1.8 | 1.8 |
| 122 | 669669.0 | 8489889.0 | 646.0 | 350.0 | 63.00 | 0.07 | 6.87 | 9.9 | 1.8 | 1.8 |
| 123 | 669762.0 | 8489935.0 | 646.0 | 350.0 | 69.00 | 0.07 | 6.87 | 6.5 | 1.8 | 1.8 |
| 124 | 669565.0 | 8489924.0 | 646.0 | 350.0 | 68.00 | 0.07 | 6.87 | 7.6 | 1.8 | 1.8 |
| 125 | 669585.0 | 8489662.0 | 646.0 | 350.0 | 43.00 | 0.07 | 6.87 | 1.3 | 1.8 | 1.8 |

Nota. En la tabla se muestra datos de entrada y salida que se utilizarán para el entrenamiento de los modelos de Machine Learning.

En la Figura 12 se muestra los puntos de ubicación de la toma de muestras de diferentes artículos recolectados para el entrenamiento de los datos.

Figura 12

Puntos de muestra de nuestros datos.



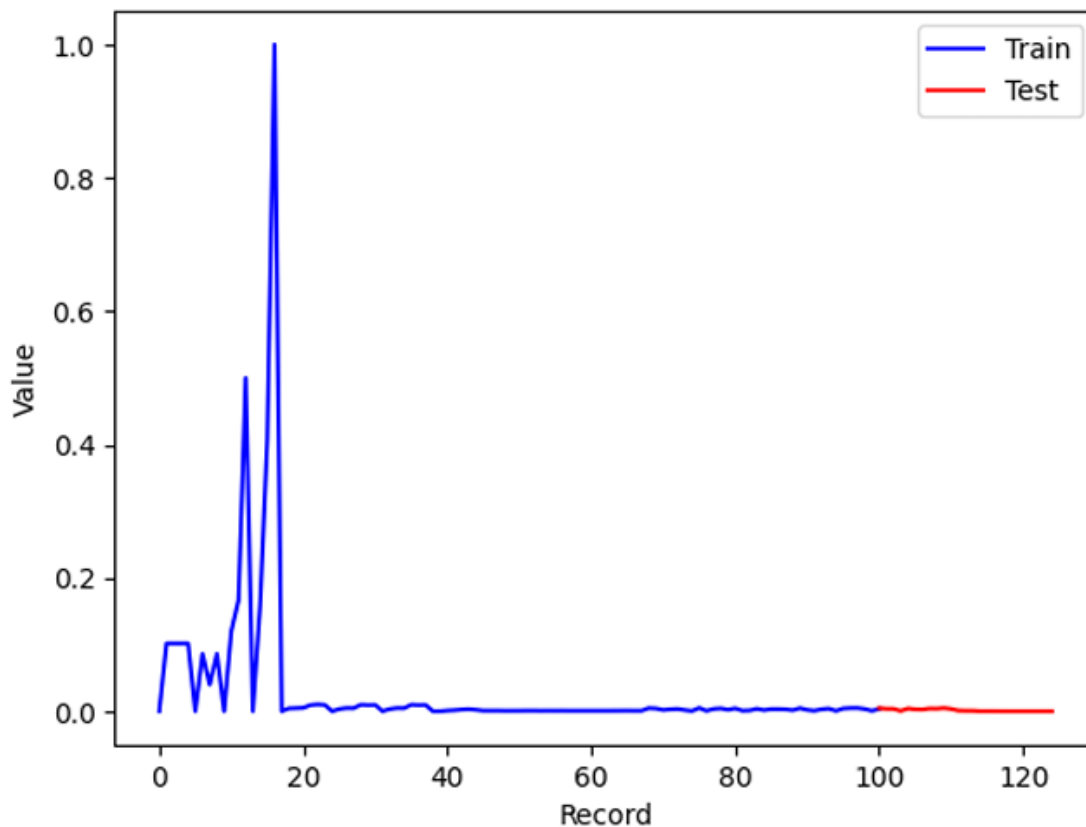
Nota. Puntos de localización de los ensayos encontrados en artículos.

B. Procesamiento de Datos

En esta parte del flujo del trabajo se realizó el preprocesamiento y procesamiento de los datos que significa ello, que en el conjunto de datos que hemos recopilados existían datos faltantes algunos estudios de laboratorio no tenían parámetros iguales a los demás entonces lo que realizamos fue rellenar datos faltantes aplicando interpolación de datos y algunos ingresando el promedio de los datos en los faltantes. Por otra parte, realizamos el escalamiento de los datos por los valores entre un parámetro y otro eran muy grandes. También se realizó la separación de los datos en Training and Testing, el 80% para Training y 20% para Testing.

Figura 13

Datos de entrenamiento y testeo para los modelos.



Nota. La figura muestra la separación de los 126 registros encontrados en entrenamiento y testeo.

Los datos de la tabla 4 se escalaron antes del entrenamiento en los modelos aplicados, porque las magnitudes de cada característica tenían mucha diferencia en su valor. este procedimiento se muestra en la figura 14. Los datos de entrada son Temperatura, PH, Nitratos y Turbidez que nos darán como salida los coliformes totales.

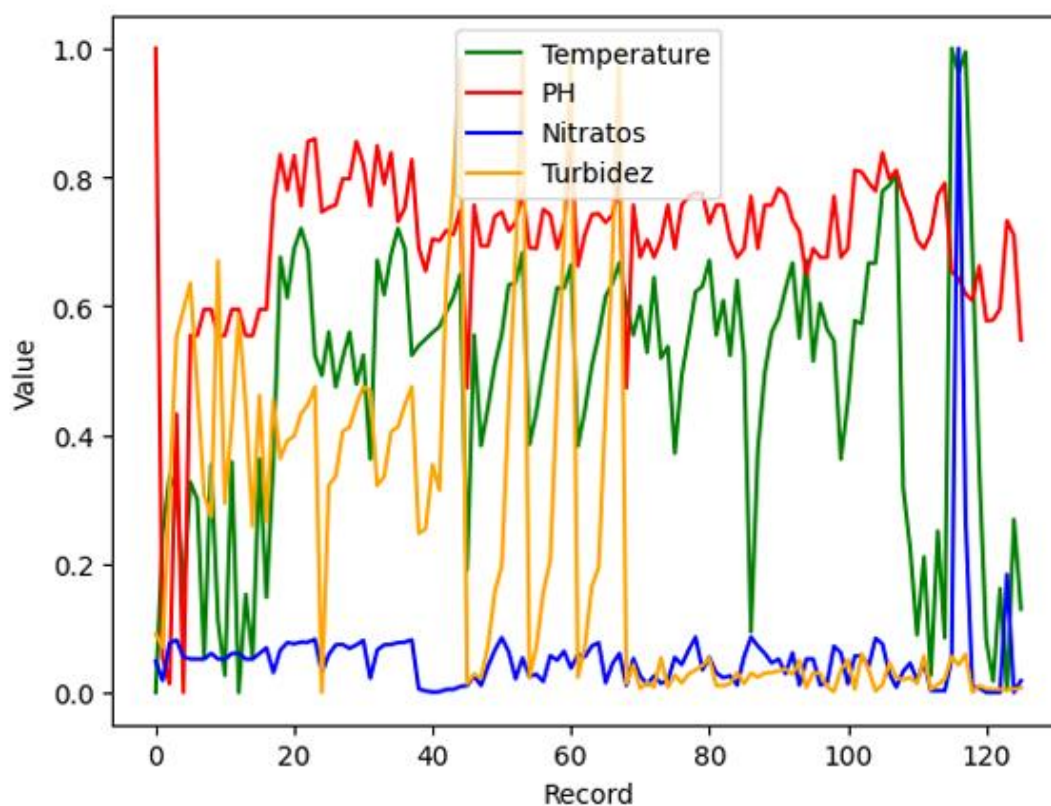
Se realiza un escalamiento de los datos en un rango de [0-1] debido a que los coliformes totales sus magnitudes eran muy elevadas alrededor de 1000 en algunos casos y el

campo de nitratos con valores alrededor de 1 entonces estos valores difieren mucho en su valor en el entrenamiento los modelos no eran capaces de predecir por lo que se realizó la técnica de escalamiento de Máximo y mínimo.

Si observan la (Figura 13), verán que los valores de pH con respecto a Nitratos tienen una diferencia, pero en referencia a los valores iniciales escalados esa diferencia se reduce.

Figura 14

Datos escalados para su entrenamiento



Nota. La figura muestra las características de los datos escalados.

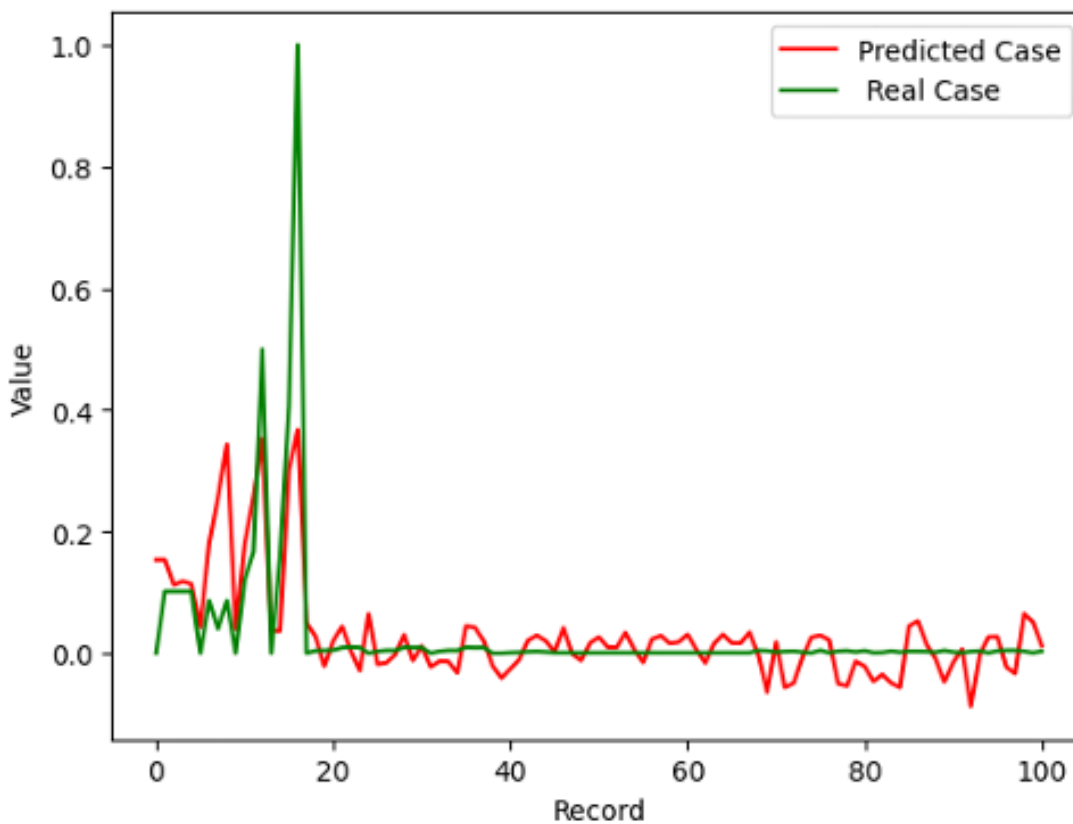
C. Aplicación e Inteligencia Artificial

a. Modelo de Regresión Lineal

Se aplicó este modelo de los cuales se tuvo buenos resultados en la predicción, los resultados de las métricas de evaluación de cada modelo se encontrarán en el ítem de predicción de coliformes totales. En este ítem se mostrará el comportamiento de los entrenamientos y el teste de cada modelo.

Figura 15

Modelo de regresión lineal



Nota. Fase de entrenamiento del modelo de regresión lineal múltiple

En la Tabla 5, se muestra los resultados de aplicación de las métricas de evaluación en el modelo de regresión lineal.

Tabla 6

Métricas en la Regresión Lineal

| Métricas | Valor |
|-----------------|--------------|
| MAE | 0.044 |
| MSE | 0.007 |
| RMSE | 0.086 |
| R ₂ | 0.480 |
| MAPE | 18.120 |
| SMAPE | 0.320 |

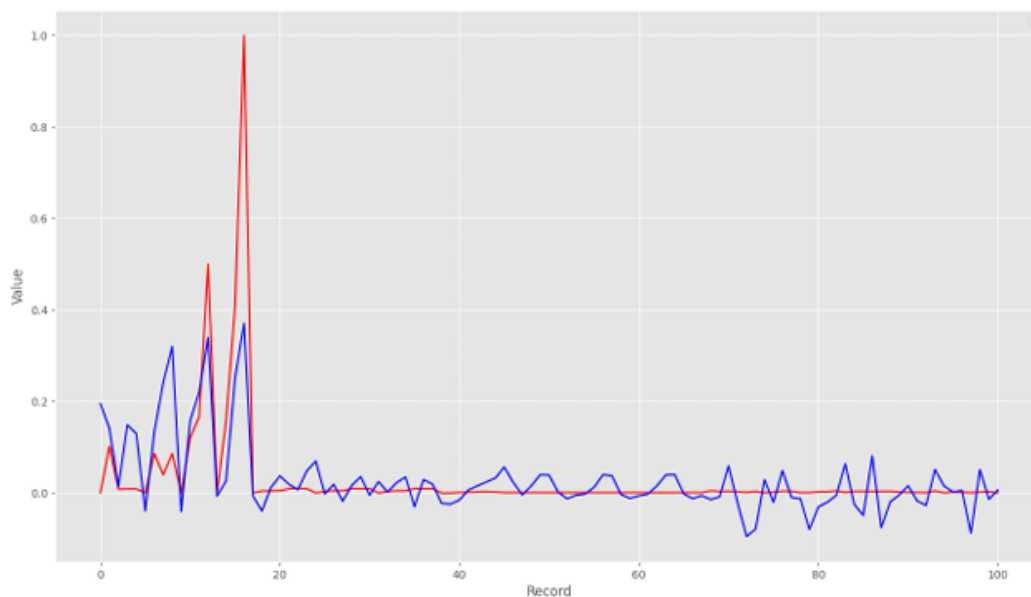
Nota. Resultados de las métricas evaluadas en el modelo de regresión lineal.

b. Predicción de coliformes totales con árboles de decisión

La predicción de coliformes totales en fuentes de agua es esencial para garantizar la calidad y seguridad del recurso hídrico. Los árboles de decisión se han utilizado como herramientas efectivas en este ámbito, permitiendo modelar y predecir la presencia de estos indicadores bacteriológicos. (Apaza, Gàrate, Cuadros y Ccasa, 2022)

En esta tesis se aplicó esta modelo para la predicción de coliformes totales de los cuales se muestra los resultados del modelo junto con los casos reales, en la Figura

Figura 16

Modelo Decision Tree

Nota. Fase de entrenamiento del modelo de árboles de decisión.

Los árboles de decisión representan una herramienta valiosa para la predicción de coliformes totales en fuentes de agua, contribuyendo a la gestión y aseguramiento de la calidad del recurso hídrico. La implementación de estos modelos puede mejorar la capacidad de respuesta ante potenciales contaminaciones y apoyar en la toma de decisiones para el tratamiento y distribución de agua potable (Segura y Cols, 2021).

En la Tabla 6, se muestra los resultados de aplicación de las métricas de evaluación en el modelo de Decision Tree.

Tabla 7

Métricas de Evaluación en Decision Tree

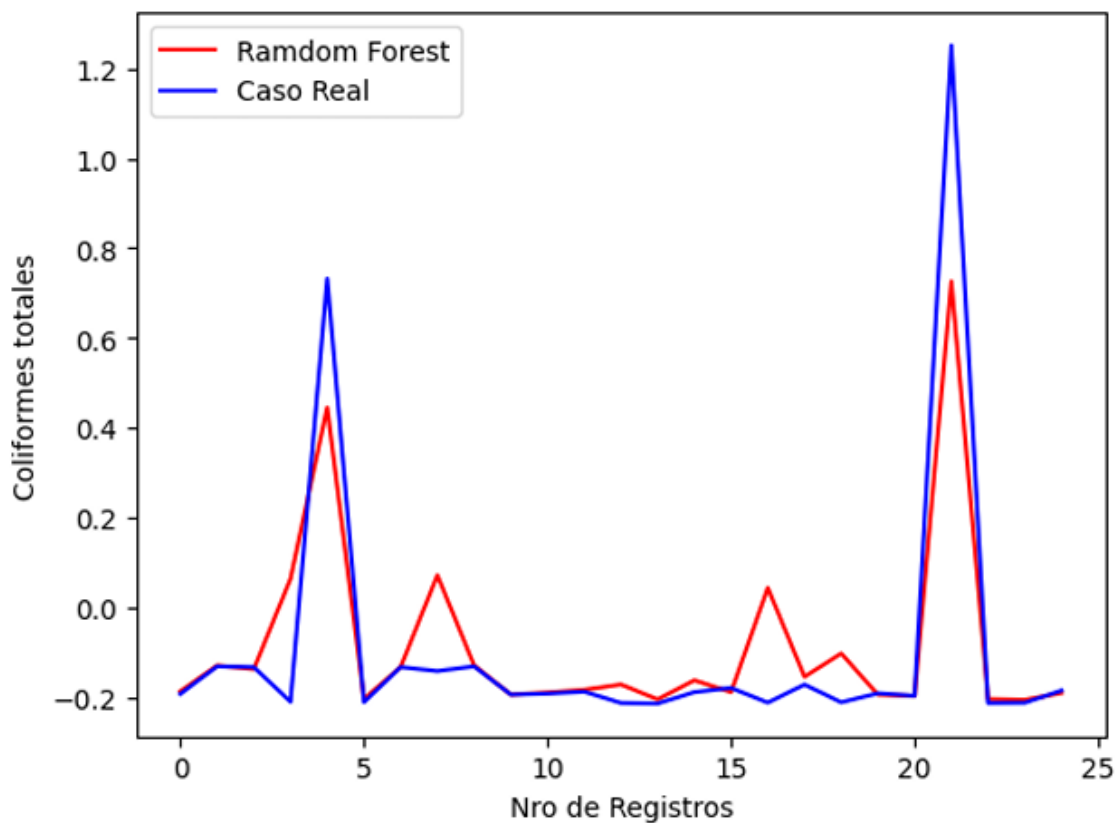
| Métricas | Valor |
|-----------------|--------------|
| MAE | 0.421 |
| MSE | 0.544 |
| RMSE | 0.738 |
| R ₂ | 0.598 |
| MAPE | 20.098 |
| SMAPE | 0.942 |

Nota. Resultados de las métricas evaluadas en el modelo de Decision Tree.

c. Predicción de coliformes totales con Random Forest

El modelo de Random Forest es una técnica de aprendizaje automático que combina múltiples árboles de decisión para mejorar la precisión y robustez en tareas de predicción. En el contexto de la calidad del agua, se ha utilizado para predecir la presencia de coliformes totales, indicadores clave de contaminación bacteriológica (Mohammeed, Hameed y Seidu, 2017).

Figura 17

Modelo Random Forest

Nota. Fase de entrenamiento del modelo de Random Forest.

El uso de Random Forest en la predicción de coliformes totales en fuentes de agua ha demostrado ser una herramienta eficaz para mejorar la gestión y monitoreo de la calidad del agua. Su capacidad para manejar datos complejos y proporcionar predicciones precisas lo convierte en una opción valiosa para las autoridades ambientales y los investigadores en el campo de la salud pública (Segura y Cols, 2021).

En la Tabla 7, se muestra los resultados de aplicación de las métricas de evaluación en el modelo de Random Forest.

Tabla 8

Métricas de Evaluación en Random Forest

| Métricas | Valor |
|-----------------|--------------|
| MAE | 0.269 |
| MSE | 0.125 |
| RMSE | 0.353 |
| R ₂ | 0.687 |
| MAPE | 18.254 |
| SMAPE | 0.745 |

Nota. Resultados de las métricas evaluadas en el modelo de Random Forest.

d. Predicción de coliformes totales con Redes Neuronales Artificiales con 1000 neuronales iniciales y cuatro capas

Las Redes Neuronales Artificiales (ANN) son modelos computacionales inspirados en el funcionamiento del cerebro humano, capaces de identificar y aprender patrones complejos a partir de datos. En el ámbito de la calidad del agua, las ANN se han empleado para predecir la presencia de coliformes totales, indicadores clave de contaminación microbiológica (Quiñones, Ochoa y Gamarra, 2020).

En esta investigación usamos 4 capas de redes neuronales iniciando la primera capa con 1000 neuronas y degradando en las siguientes capas a mitad, así como se muestra en la Tabla 9.

Tabla 9

Primer Modelo de Red Neuronal Artificial

| Layer (type) | Output Shape | Param # |
|--------------------------|-----------------------|----------------|
| dense_4 (Dense) | (None, 1000) | 10,000 |

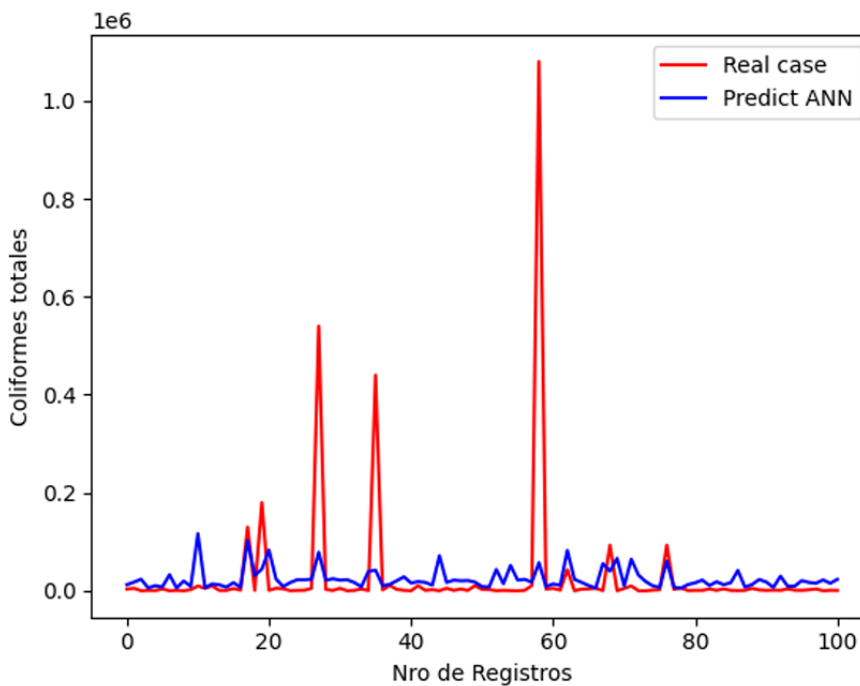
| Layer (type) | Output Shape | Param # |
|-----------------|--------------|---------|
| dense_5 (Dense) | (None, 500) | 500,500 |
| dense_6 (Dense) | (None, 250) | 125,250 |
| dense_7 (Dense) | (None, 1) | 251 |

Nota. En la tabla se muestra datos del modelo propuesto para la predicción de Coliformes totales con ANN.

El modelo propuesto predijo los coliformes totales como se muestra en la Figura 18, aplicando el modelo con 4 capas e iniciando con 1000 neuronas en la primera capa y en cada siguiente capa reduciendo a la mitad.

Figura 18

Primer Modelo de ANN

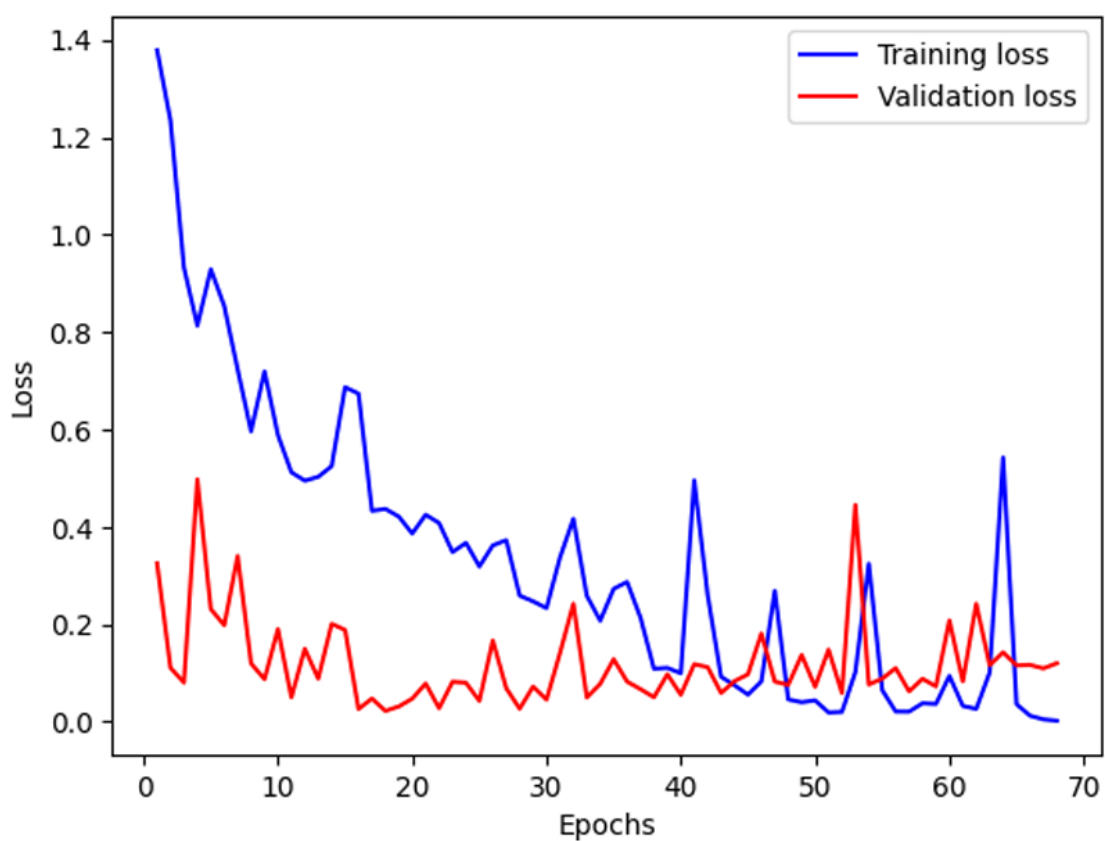


Nota. Fase de entrenamiento del modelo de la primer Red Neuronal Artificial.

Se aplicó las métricas de pérdida de error en el primer modelo de red neuronal y se obtuvo la pérdida como se muestra en la Figura 19, donde va bajando desde el valor de 1.75 hasta el rango entre 1.25 hasta 1.00 en los datos del testeo del modelo.

Figura 19

Primer Modelo de ANN - Validation Loss



Nota. Fase de entrenamiento del modelo de la primer Red Neuronal Artificial.

e. Predicción de coliformes totales con Redes Neuronales Artificiales con 1000 neuronales iniciales y 7 capas

Las Redes Neuronales Artificiales (ANN) se han aplicado en diversos estudios para predecir la presencia de coliformes totales en cuerpos de agua, en esta tesis se especifica de una ANN con 100 neuronas y 7 capas en el modelo creado, en la Tabla 10, se muestra el esquema de cómo está nuestro segundo modelo de red neuronal artificial.

Tabla 10

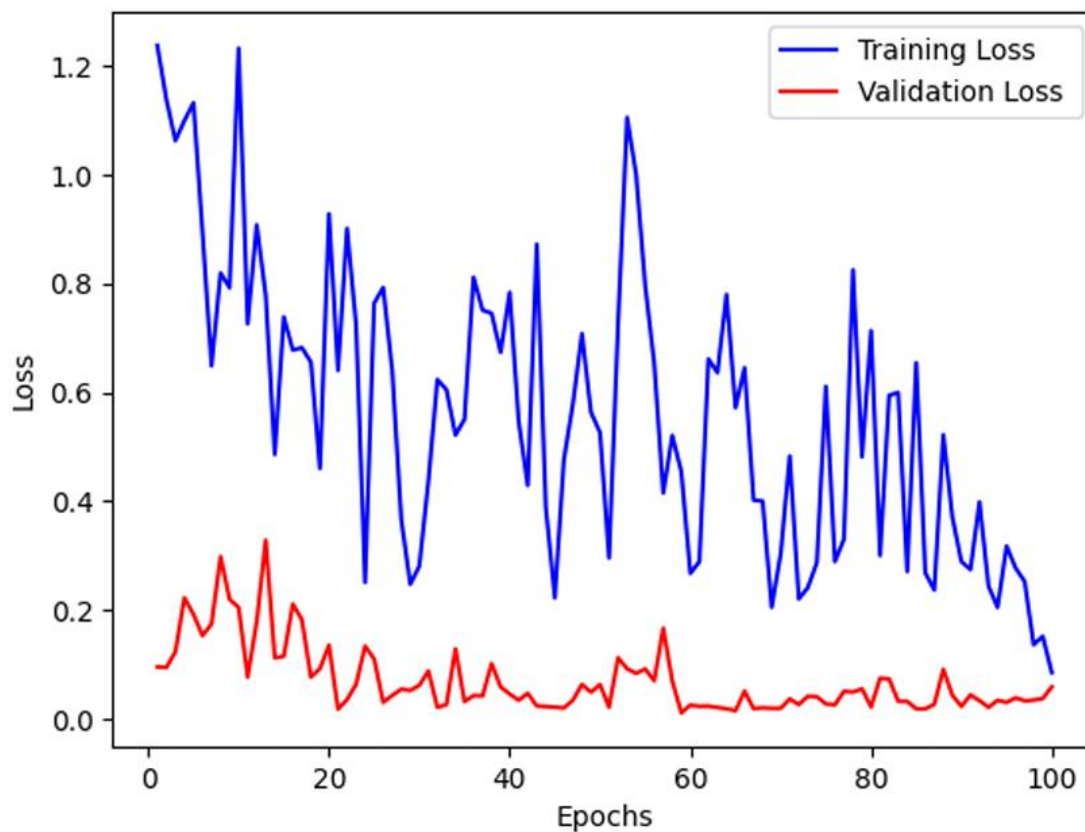
Segundo Modelo de ANN

| Layer (type) | Output Shape | Param # |
|---------------------|---------------------|----------------|
| dense_52 (Dense) | (None, 1000) | 10,000 |
| dropout (Dropout) | (None, 1000) | 0 |
| dense_53 (Dense) | (None, 500) | 500,000 |
| Dropout_1 (Dropout) | (None, 500) | 0 |
| dense_54 (Dense) | (None, 250) | 125,250 |
| Dropout_2 (Dropout) | (None, 250) | 0 |
| dense_55 (Dense) | (None, 1) | 251 |

Nota. Fase de entrenamiento del modelo de la segunda Red Neuronal Artificial.

Se aplicó las métricas de pérdida de error en el segundo modelo de red neuronal y se obtuvo la pérdida como se muestra en la figura 20, donde va bajando desde el valor de 1.20 hasta el rango entre 1.2 hasta 1.00 en los datos del testeo del modelo.

Figura 20

Segundo Modelo de ANN - Validation Loss

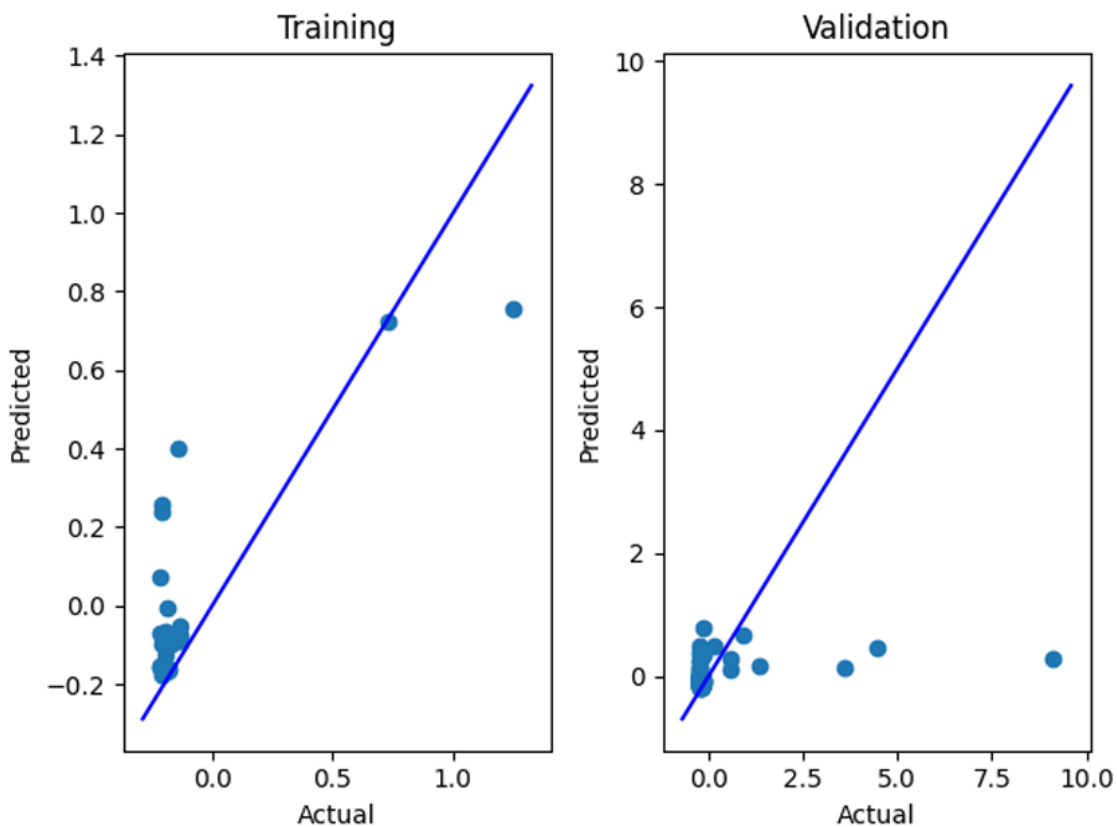
Nota. Fase de entrenamiento del modelo de la primer Red Neuronal Artificial.

D. Predicción de Coliformes Totales

Los modelos más adecuados que resultaron en la predicción de coliformes totales fueron los modelos de redes neuronales artificiales, estos resultados se muestran en las siguientes gráficas de regresión lineal entre los modelos de red neuronales y los casos reales, donde se obtuvo valores muy cercanos entre los datos reales y los predichos por los modelos propuestos en esta tesis.

Figura 21

Primer Modelo de ANN regresión lineal en los datos de Training y testeo



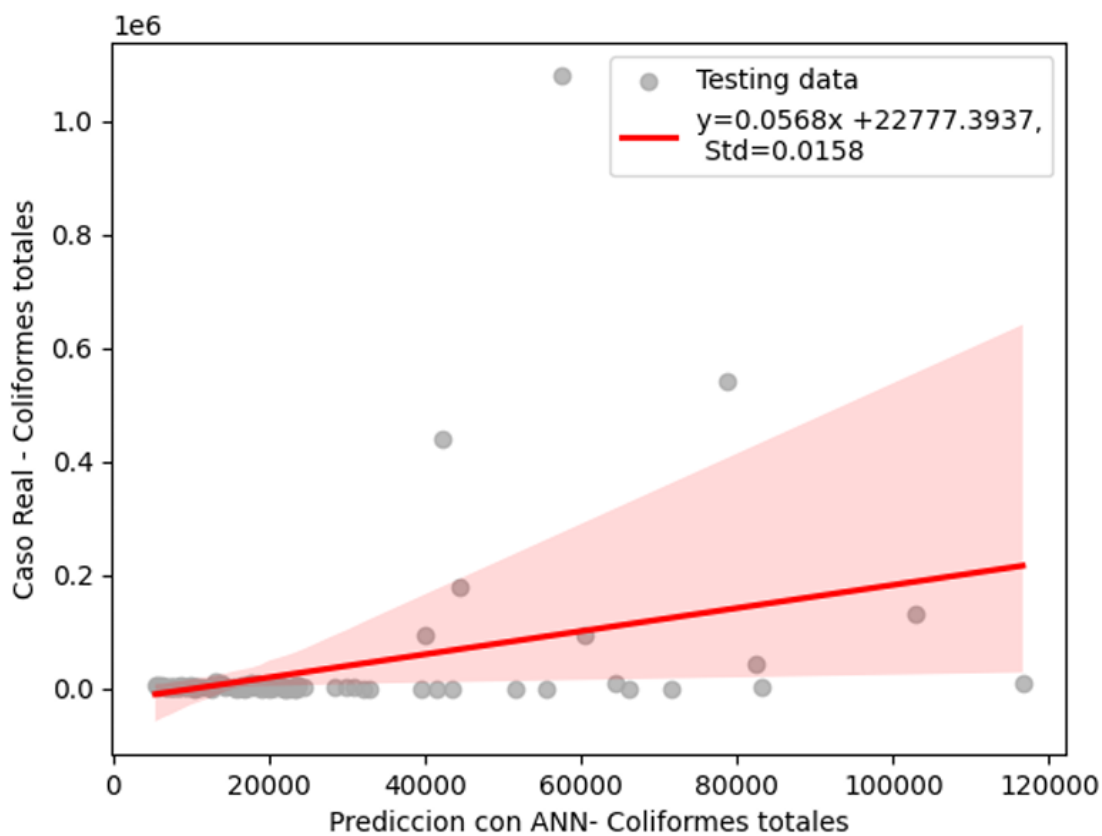
Nota. Fase de entrenamiento y testeo del modelo de la primer Red Neuronal Artificial.

En la figura 22 y 23, se muestran las regresiones lineales de los datos predichos con el primer modelo y casos reales además se muestra el segundo modelo de red neuronal con los casos reales en la predicción de coliformes totales en la calidad de agua. Según las gráficas el primer modelo de regresión lineal tiene una desviación estándar de 0.0158 que es un valor mínimo en la diferencia de los datos reales y los predichos por el modelo 1, y en el segundo modelo el valor de la desviación estándar es de 0.0099 que es un valor más

pequeño que del primero modelo esto quiere decir que el modelo número 1, tiene mayor capacidad de predecir coliformes totales utilizando esa arquitectura de red neuronal.

Figura 22

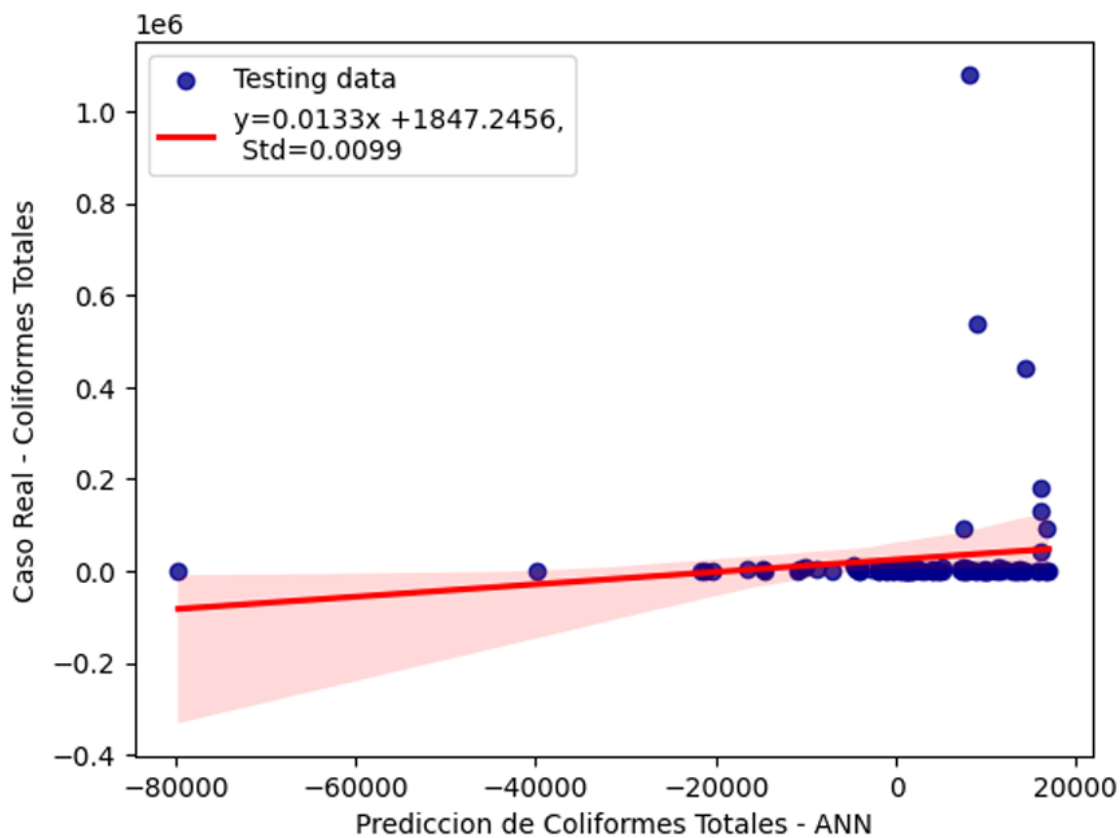
Primer Modelo de ANN regresión lineal en los datos reales y predichos.



Nota. Regresión lineal entre el modelo de la primer Red Neuronal Artificial y los datos reales.

Figura 23

Segundo Modelo de ANN regresión lineal en los datos reales y predichos.



Nota. Regresión lineal entre el modelo del segundo modelo de Red Neuronal Artificial y los datos reales.

4.6 Análisis de datos

4.6.1 Procedimiento Estadístico

El procedimiento estadístico a utilizar será:

Técnicas estadísticas: Método T-Student

A. Hipótesis estadísticas: Hipótesis General

Aplicando la Inteligencia Artificial si se predicen agentes patológicos mediante parámetros de la calidad del agua en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

✓ Datos

Tabla 11

Datos de entrada para los modelos de aprendizaje

| Nº | Este | Norte | Cond. Eléctrica | Dureza Total | DBO | Nitratos | pH | T° | Turbidez |
|-----|----------|-----------|-----------------|--------------|--------|----------|-------|-----|----------|
| 0 | 682457.0 | 8483792.0 | 251.0 | 132.0 | 6.90 | 5.60 | 10.00 | 1.6 | 14.00 |
| 1 | 676659.0 | 8489942.0 | 353.0 | 259.0 | 117.20 | 2.20 | 3.00 | 7.1 | 11.00 |
| 2 | 673857.0 | 8489386.0 | 214.0 | 89.0 | 65.25 | 8.80 | 2.70 | 9.1 | 46.00 |
| 3 | 669719.0 | 8489597.0 | 356.0 | 293.0 | 106.48 | 9.40 | 5.80 | 8.6 | 80.00 |
| 4 | 667376.0 | 8491705.0 | 170.0 | 293.0 | 39.00 | 6.40 | 2.60 | 5.0 | 86.00 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 121 | 669632.0 | 8489984.0 | 648.0 | 360.0 | 43.00 | 0.08 | 6.88 | 2.0 | 1.70 |
| 122 | 669669.0 | 8489889.0 | 680.0 | 330.0 | 28.00 | 0.09 | 7.01 | 5.2 | 1.65 |
| 123 | 669762.0 | 8489935.0 | 720.0 | 370.0 | 31.00 | 21.05 | 8.02 | 1.7 | 1.55 |
| 124 | 669565.0 | 8489924.0 | 620.0 | 380.0 | 40.00 | 0.09 | 7.85 | 7.6 | 1.74 |
| 125 | 669585.0 | 8489662.0 | 610.0 | 270.0 | 27.00 | 2.15 | 6.65 | 4.5 | 2.01 |

Nota. En la tabla se muestra datos de entrada que son algunas características tomadas en laboratorio por estudios de otros artículos.

Tabla 12

Datos de entrada para los modelos de aprendizaje

| Nº | Coliformes Totales |
|----|--------------------|
| 0 | 300.0 |
| 1 | 110000.0 |
| 2 | 9000.0 |
| 3 | 9780.0 |
| 4 | 9752.0 |

| Nº | Coliformes Totales |
|------------|--------------------|
| ... | ... |
| 121 | 423.0 |
| 122 | 412.0 |
| 123 | 560.0 |
| 124 | 750.0 |
| 125 | 880.0 |

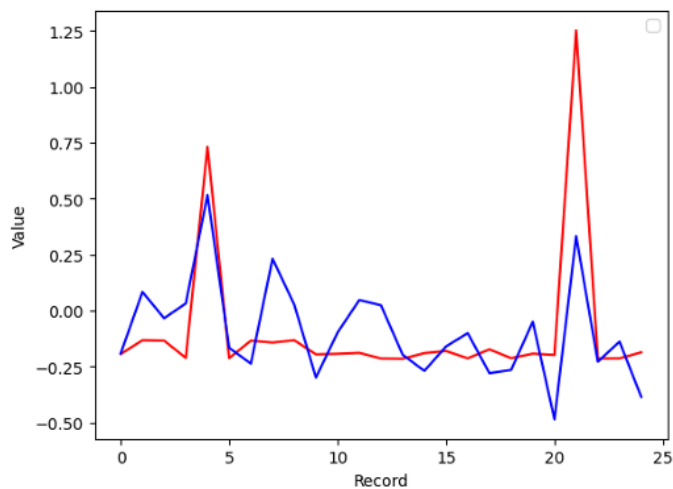
Nota. En la tabla se muestra datos de entrada que son algunas características tomadas en laboratorio por estudios de otros artículos.

✓ **Predicción de los Coliformes totales con los diferentes modelos de Machine Learning.**

Para realizar la predicción de los coliformes totales, se evaluó los modelos desarrollados con los datos del testeo que se separó al inicio de realizar la distribución y separación de los datos totales. Todos los resultados se muestran en las siguientes figuras 24, 25, 26 por cada modelo aplicado en esta tesis.

Figura 24

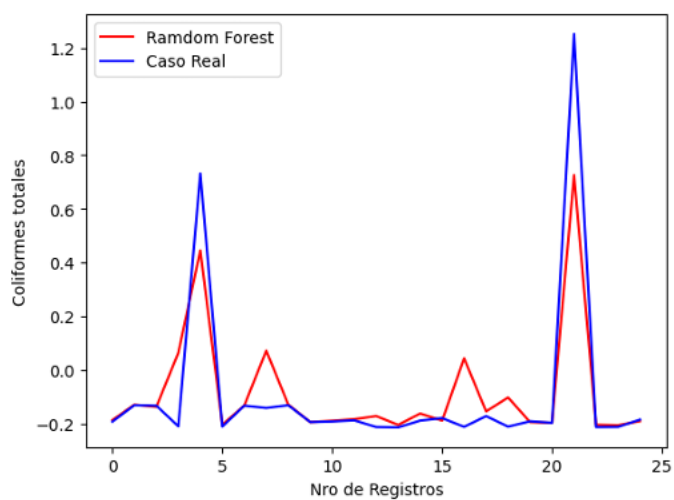
Predicción de coliformes totales mediante regresión lineal.



Nota. Predicción de coliformes totales mediante el modelo Regresión lineal en los datos de testo.

Figura 25

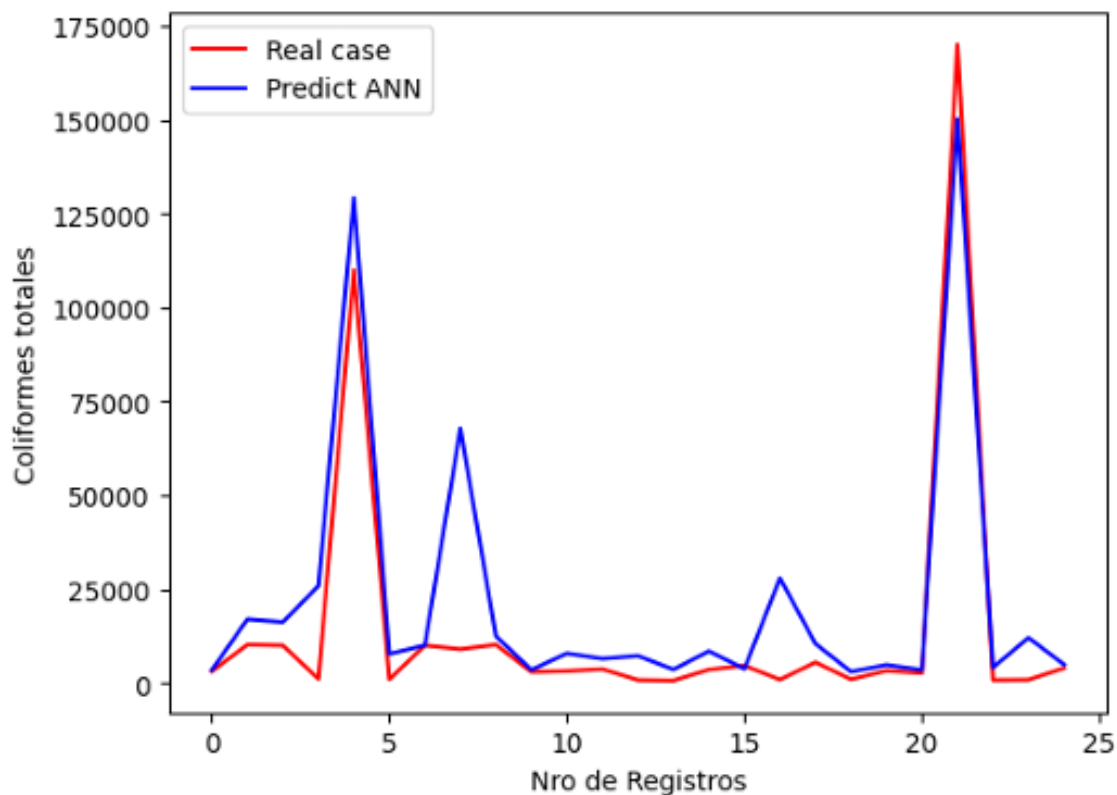
Predicción de coliformes totales mediante Random Forest



Nota. Predicción de coliformes totales mediante el modelo Random Forest en los datos de testo.

Figura 26

Predicción de coliformes totales mediante Redes Neuronales Artificiales



Nota. Predicción de coliformes totales mediante el modelo de Redes Neuronales en los datos de testo.

B. Hipótesis estadística: Hipótesis Específica 1

La correlación entre los valores reales de los parámetros microbiológicos y parasitológicos de la calidad del agua y los valores predichos por la aplicación de la inteligencia artificial presenta un grado de correlación $R^2 > 50\%$ en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

✓ **Datos de testeo**

Tabla 13

Resultados de la aplicación del modelo de Red Neuronal

| Caso Real | ANN | Caso Real | ANN |
|------------------|------------|------------------|------------|
| 3111 | 11712.211 | 10200 | 34422.64 |
| 10000 | 33561.656 | 1100 | 66033.83 |
| 110000 | 125692.980 | 987 | 23528.00 |
| 10000 | 18310.934 | 9000 | 90761.35 |
| 10200 | 28451.133 | 2936 | 11417.082 |
| 3199 | 26537.604 | 3696 | 23645.86 |
| 790 | 44440.086 | 646 | 11538.451 |
| 3564 | 41027.406 | 4638 | 11036.572 |
| 880 | 50058.676 | 5500 | 17684.518 |
| 962 | 28048.693 | 3270 | 13760.779 |
| 2622 | 10725.924 | 170000 | 172392.70 |
| 781 | 11535.926 | 851 | 20139.646 |
| 3964 | 12716.375 | ... | ... |

Nota. Aplicación de la Red Neuronales en la predicción de coliformes totales una comparación con los datos reales.

✓ **Hipótesis estadísticas**

$$H_0: R^2 < 50\% \quad \text{-----} \rightarrow \text{Ec. 7}$$

$$H_1: R^2 > 50\% \quad \text{-----} \rightarrow \text{Ec. 8}$$

Donde:

r^2 : Grado de correlación de los resultados entre los Parámetros patológicos encontrados en laboratorio y los Parámetros patológicos predichos por la Inteligencia Artificial.

✓ **Estadístico**

Tabla 14

Datos Estadísticos O.E.1

| Variable | Media | Desv. Est. | Varianza | Coef. Var. | Mínimo | Mediana | Máximo | Rango |
|-----------------|--------------|-------------------|-----------------|-------------------|---------------|----------------|---------------|--------------|
| Caso Real | 14916 | 38769 | 1503049556 | 259.92 | 646 | 3270 | 170000 | 169354 |
| ANN | 7567 | 39210 | 1537405610 | 104.37 | 10726 | 23646 | 172393 | 161667 |

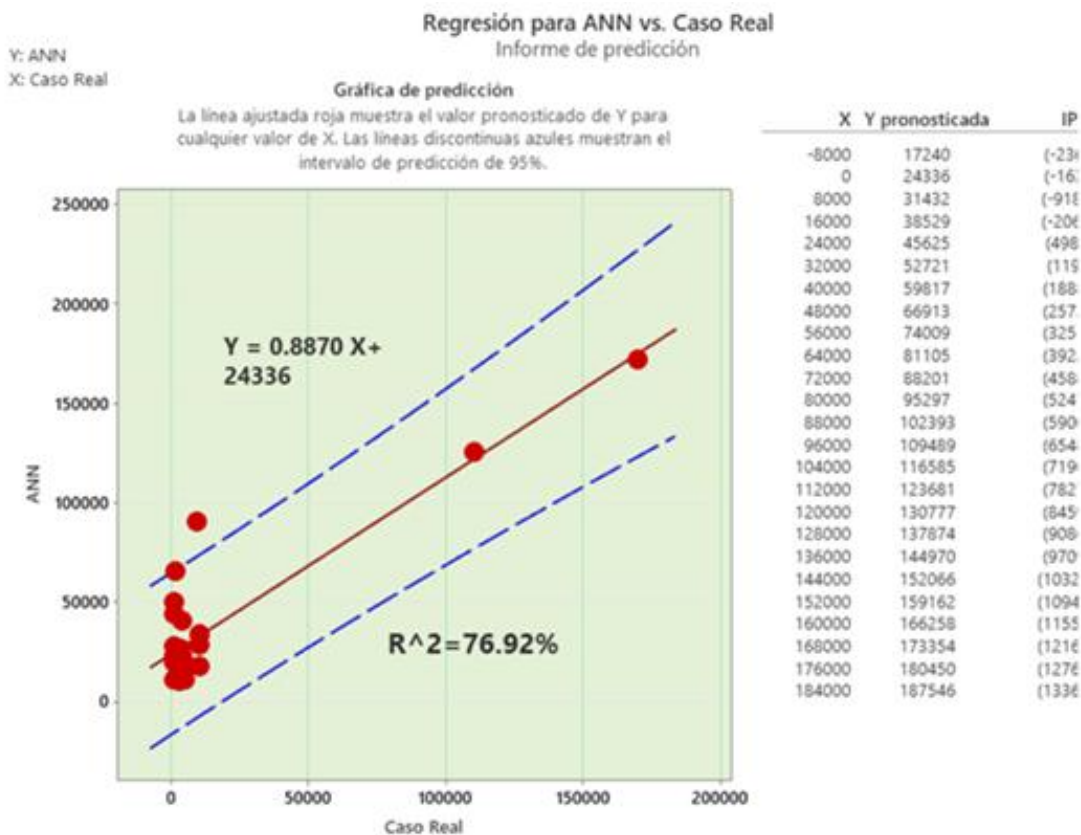
Nota. Resultados Estadísticos del Hipótesis Específico 1.

✓ **Nivel de significancia**

$$\alpha = 5\% \cong 0,05$$

-----> *Ec. 9*

Figura 27

Regresión lineal cálculo de R^2 

Nota. Regresión lineal entre los datos reales y los resultados de la aplicación del modelo de redes neuronales artificiales.

✓ Interpretación

Se realizó la comparación de los datos generados por las redes neuronales artificiales y los datos reales en un modelo de regresión lineal para comparar la diferencia entre estos dos resultados mediante el programa Minitab, en donde el R^2 nos dio como resultados de 76,92% quiere decir que al 95% de seguridad la correlación de los resultados tiene una aproximación mayor al 50%, esto quiere decir que aceptamos la hipótesis alterna lo que significa que el uso

de laboratorio y la Inteligencia Artificial en parámetros de la calidad del agua se obtuvo un grado de correlación $R^2 > 50\%$ en la determinación de los agentes patológicos. esto indica que las redes neuronales pueden predecir coliformes totales con una seguridad del 95%.

C. Hipótesis estadísticas: Hipótesis Específica 2

Realizando pruebas de entrenamiento y testeo se determinó que el modelo de inteligencia artificial Neuronal Network presenta mayor precisión en la predicción de los parámetros microbiológicos y parasitológicos con respecto al Machine Learning Regression en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

✓ Datos de testeo aprendizaje supervisado y el modelo de redes neuronales artificiales

En el análisis de los datos se tomaron de los 126 datos 80% para el entrenamiento y el 20% para el testeo. En la tabla 14, se muestra el resultado de la aplicación del modelo de regresión lineal y los datos reales para así realizar una regresión y obtener el grado de separación de los datos reales con los datos del modelo generado en la predicción de coliformes totales.

Por otro lado, en la tabla 15, se tiene los datos aplicados en la predicción de coliformes totales con el modelo de redes neuronales artificiales.

Tabla 15

Resultados de la aplicación del modelo de Regresión Lineal

| Caso Real | ANN | Caso Real | ANN |
|------------------|------------|------------------|------------|
| 300.00 | 410.19 | ... | ... |
| 110000.00 | 102807.12 | 2949.00 | 2726.98 |
| 900.00 | 9930.62 | 4638.00 | 4742.84 |
| 9780.00 | 97756.88 | 1330.00 | 1230.46 |
| 9752.00 | 96250.69 | 3777.00 | 36751.69 |
| 240.00 | 238.99 | 3801.00 | 3721.54 |
| 93000.00 | 92889.07 | 3564.00 | 3181.38 |
| 43000.00 | 41195.13 | 3696.00 | 3941.01 |
| 93000.00 | 92900.94 | 3716.00 | 3213.13 |
| 170.00 | 165.55 | 2473.00 | 2586.49 |
| 130000.00 | 108225.52 | 1319.00 | 13157.93 |
| 180000.00 | 168374.46 | 1277.00 | 12825.85 |
| 540000.00 | 539702.81 | 781.00 | 780.36 |
| 180.00 | 197.26 | 4732.00 | 5547.26 |
| 170000.00 | 166025.10 | 530.00 | 564.82 |
| 440000.00 | 420712.63 | 2242.00 | 2206.24 |
| 1080000.00 | 1059979.88 | 1778.00 | 1779.31 |
| 150.00 | 120.90 | 813.00 | 745.94 |
| 4500.00 | 4280.93 | 2126.00 | 2103.98 |
| 5000.00 | 13202.45 | 3199.00 | 4068.33 |
| 5500.00 | 5044.70 | 528.00 | 7073.39 |

Nota. Aplicación del aprendizaje supervisado con el modelo de regresión lineal en la predicción de coliformes totales una comparación con los datos reales.

Tabla 16

Resultados de la aplicación del modelo de redes neuronales artificiales

| Caso Real | ANN | Caso Real | ANN |
|------------------|------------|------------------|------------|
| 25490.86 | 24941.58 | ... | ... |
| 37224.31 | 36876.57 | 25774.19 | 25369.53 |
| 26421.31 | 26380.01 | 29954.85 | 25594.72 |
| 26504.83 | 26455.58 | 25601.03 | 25527.78 |
| 26501.83 | 26492.00 | 25862.76 | 25654.79 |
| 25484.44 | 25290.64 | 25865.32 | 25595.10 |
| 35406.00 | 35363.42 | 25839.97 | 25776.22 |
| 30058.03 | 30047.03 | 25854.09 | 25247.76 |
| 35406.00 | 35365.09 | 25856.23 | 25642.28 |
| 25476.95 | 25476.44 | 25723.28 | 25421.88 |
| 39363.50 | 39296.83 | 25599.85 | 25439.37 |
| 44711.48 | 44710.18 | 25595.36 | 25458.59 |
| 20216.89 | 20921.62 | 25542.31 | 25422.99 |
| 25478.02 | 25370.77 | 25964.90 | 25357.87 |

| Caso Real | ANN | Caso Real | ANN |
|------------------|------------|------------------|------------|
| 43641.88 | 43521.43 | 25515.46 | 25582.01 |
| 72520.94 | 72425.53 | 25698.57 | 25583.01 |
| 26975.02 | 26982.16 | 25648.94 | 25138.30 |
| 25474.81 | 25472.81 | 25545.73 | 25352.16 |
| 25940.09 | 25930.80 | 25686.17 | 25437.44 |
| 25993.57 | 25961.37 | 25800.93 | 25795.81 |
| 26047.05 | 26044.69 | 25515.24 | 2585.11 |

Nota. Aplicación del modelo de redes neuronales en la predicción de coliformes totales una comparación con los datos reales.

✓ **Hipótesis estadísticas**

$H_0: r_1^2 < r_2^2$ [El grado de distribución r_1^2 del modelo de regresión lineal es menor que el grado de distribución r_2^2 del modelo de redes neuronales]

$H_1: r_1^2 > r_2^2$ [El grado de distribución r_1^2 del modelo de regresión lineal es mayor que el grado de distribución r_2^2 del modelo de redes neuronales]

✓ **Estadístico**

En la Tabla 16, se muestra los resultados estadísticos aplicados a los datos predichos por el modelo de aprendizaje supervisado de regresión lineal y de la aplicación del modelo de redes neuronales artificiales, se muestra los resultados de la median, desviación estándar, varianza coeficiente de variación, mínimo valor, mediana y máximo valor.

Tabla 17

Datos Estadísticos O.E.2

| Variable | Media | Desv. Est. | Varianza | Coef. Var. | Mínimo | Mediana | Máximo |
|-----------------|--------------|-------------------|-----------------|-------------------|---------------|----------------|---------------|
| ANN | 25530 | 6833 | 466966474 | 26.77 | 2585 | 25459 | 72426 |
| R.L. | 34402 | 126378 | 15971449328 | 367.36 | 15 | 3343 | 1059980 |

Nota. Resultados Estadísticos del Hipótesis Específico 2.

✓ **Nivel de significancia**

$$\alpha = 5\% \cong 0,05 \quad \text{-----}\rightarrow \text{Ec. 10}$$

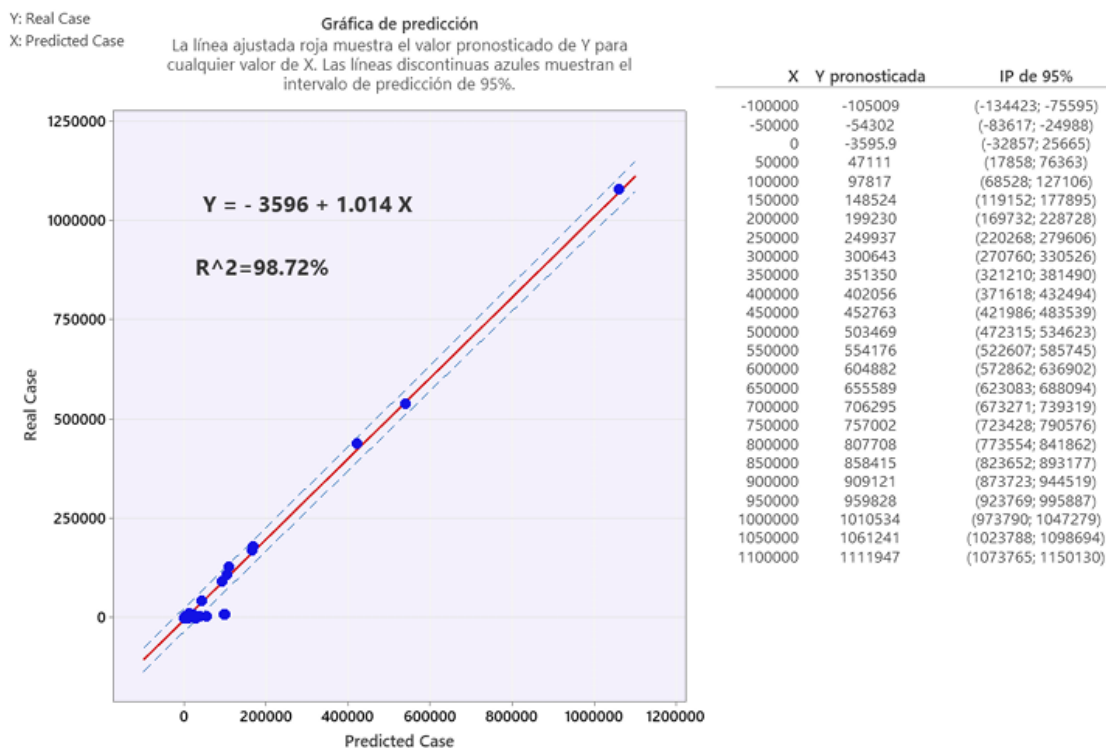
Este valor nos indicara que con el 95% se seguridad afirmaremos nuestra hipótesis alterna.

✓ **Resultados**

En la Figura 28 y 29 se muestran la ecuación que rige el comportamiento lineal de los datos entre los modelos de aprendizaje supervisado, las redes neuronales y los datos reales en cada caso, también se muestra los valores de R^2 que es un valor importante que nos permitió calcular el porcentaje de similitud entre los resultados de los modelos y nuestros datos reales así poder indicar cuál de los modelos trabaja mejor en la predicción.

Figura 28

Regresión lineal del modelo de aprendizaje supervisado cálculo de R^2



Nota. Regresión lineal entre los datos reales y los resultados de la aplicación del modelo de regresión lineal

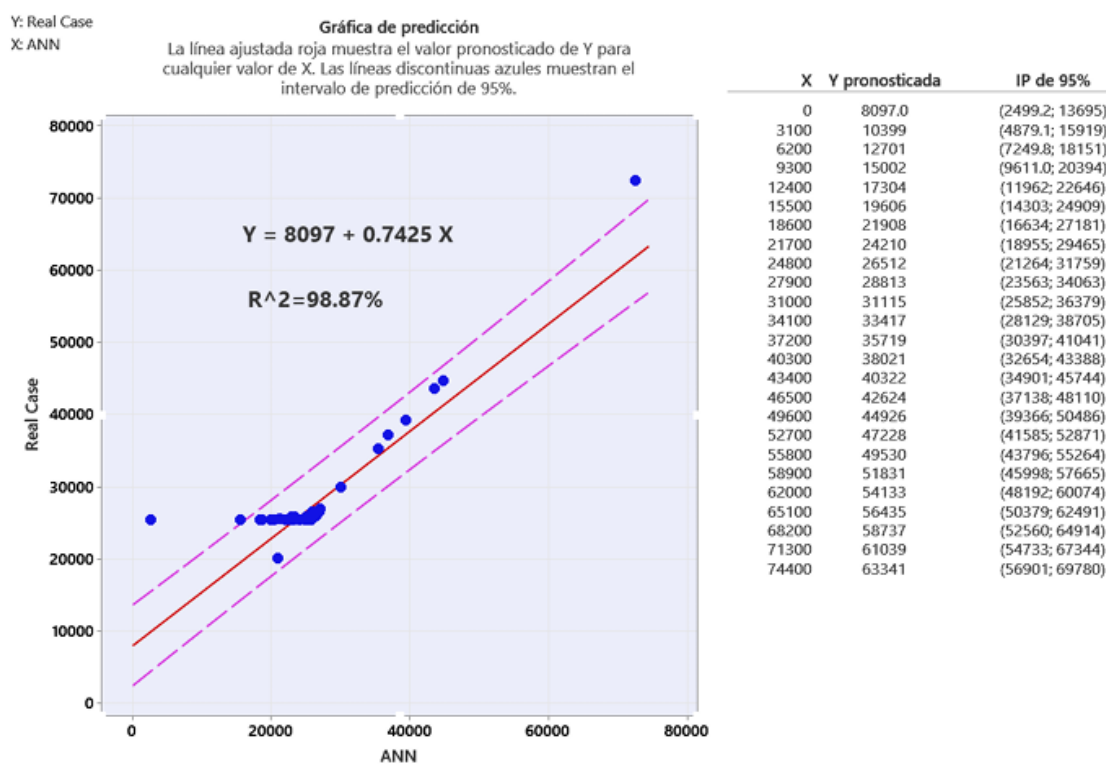
En la figura 28, nos muestra una regresión lineal que hace la comparación de los datos reales y el modelo supervisado de regresión lineal y el resultado de $R^2 = 98,72\%$ esto nos indica que el modelo supervisado predice valores de coliformes totales con una seguridad del 95% podemos afirmar que los datos se encuentran muy similares en un 98.72% entre sus valores.

En la figura 29, nos muestra una regresión lineal que hace la comparación de los datos reales y el modelo de red neuronal artificial y el resultado de $R^2 = 98,87\%$ esto nos indica

que el modelo de red neuronal artificial predice valores de coliformes totales con una seguridad del 95% podemos afirmar que los datos se encuentran muy similares en un 98.87% entre sus valores.

Figura 29

Regresión lineal del modelo de redes neuronales artificiales, cálculo de R^2



Nota. Regresión lineal entre los datos reales y los resultados de la aplicación del modelo de redes neuronales artificiales.

✓ Interpretación

Se realizó la comparación de los datos generados por las redes neuronales artificiales, aprendizaje supervisado y los datos reales en dos modelos de regresión lineal, cada modelo se

comparó la diferencia entre los valores de R^2 mediante el programa Minitab, donde el R_1^2 nos dio como resultados de 98,72% afirmamos con 95% de seguridad que los valores del modelo de aprendizaje supervisado y los datos reales tiene una afinidad de ese porcentaje. Por otro lado, el valor de R_2^2 resulto de 98,87% afirmamos con 95% de seguridad que el modelo de redes neuronales artificiales y los datos reales se aproximan en ese porcentaje de dispersión, entonces con estos dos valores podemos afirmar con una seguridad del 95%, el grado de distribución r_1^2 del modelo de regresión lineal es mayor que el grado de distribución r_2^2 del modelo de redes neuronales. Esto significa que las redes neuronales artificiales predicen mejor que los modelos de aprendizaje supervisado.

D. Hipótesis estadísticas: Hipótesis Específica 3

Para realizar las pruebas de entrenamiento y testeo fueron indispensables los parámetros físicos químicos del agua los cuales se aplicaron como datos de entrada en los modelos de inteligencia artificial para determinar la predicción de los parámetros microbiológicos y parasitológicos en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

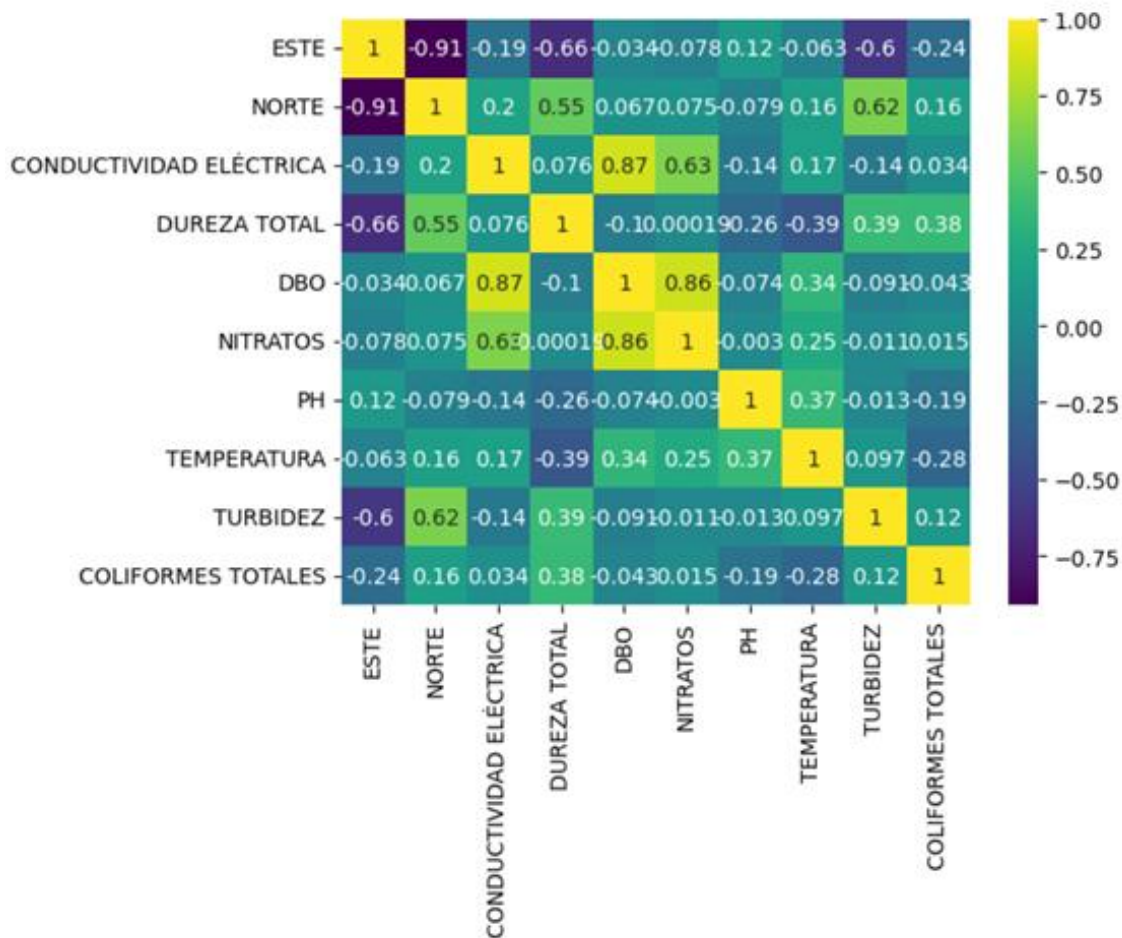
V. Resultados y Discusión

5.1 Resultados para el objetivo general

Aplicando la Inteligencia Artificial si se predicen agentes patológicos mediante parámetros de la calidad del agua en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

Figura 30

Matriz de confusión de los parámetros de la calidad del agua.



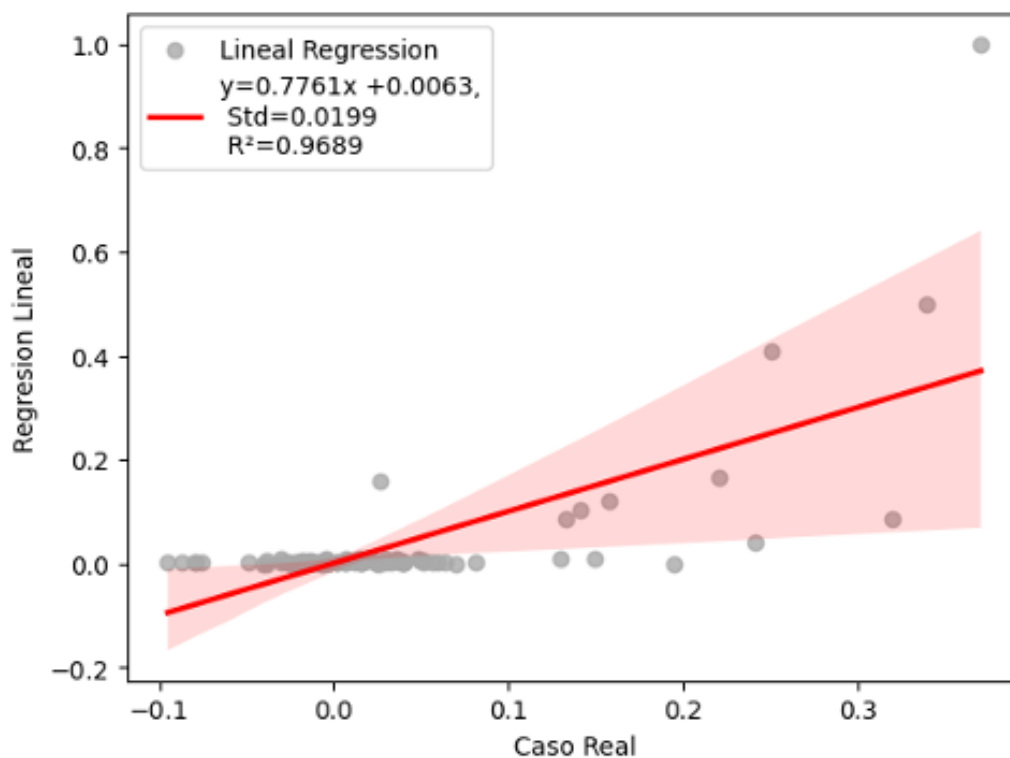
Nota. En esta figura se muestra los parámetros de entrada y salida de los datos de la calidad del agua.

En la Figura 30, se muestra 10 paramétricos de la calidad de agua. Los parámetros de entrada fueron (Este, Norte, Conductividad eléctrica, Dureza Total, DBO, Nitrato, H, Temperatura, y Turbidez) y como parámetros de salida (Coliformes totales). con los parámetros de entrada se predijo los parámetros de salida que es la predicción de agentes patológicos. Para los resultados del objetivo general que es la predicción de agentes patológicos se aplicó modelos de aprendizaje supervisado (Decision, Random forest) y redes neuronales artificiales.

En la figura 31, se muestra la regresión lineal de los modelos predictivo de Regresión Lineal. con un ajuste de $R^2 = 0,9689$, quiere decir que predice con un porcentaje igual a 96,89%.

Figura 31

Resultado del Modelo de Regresión Lineal

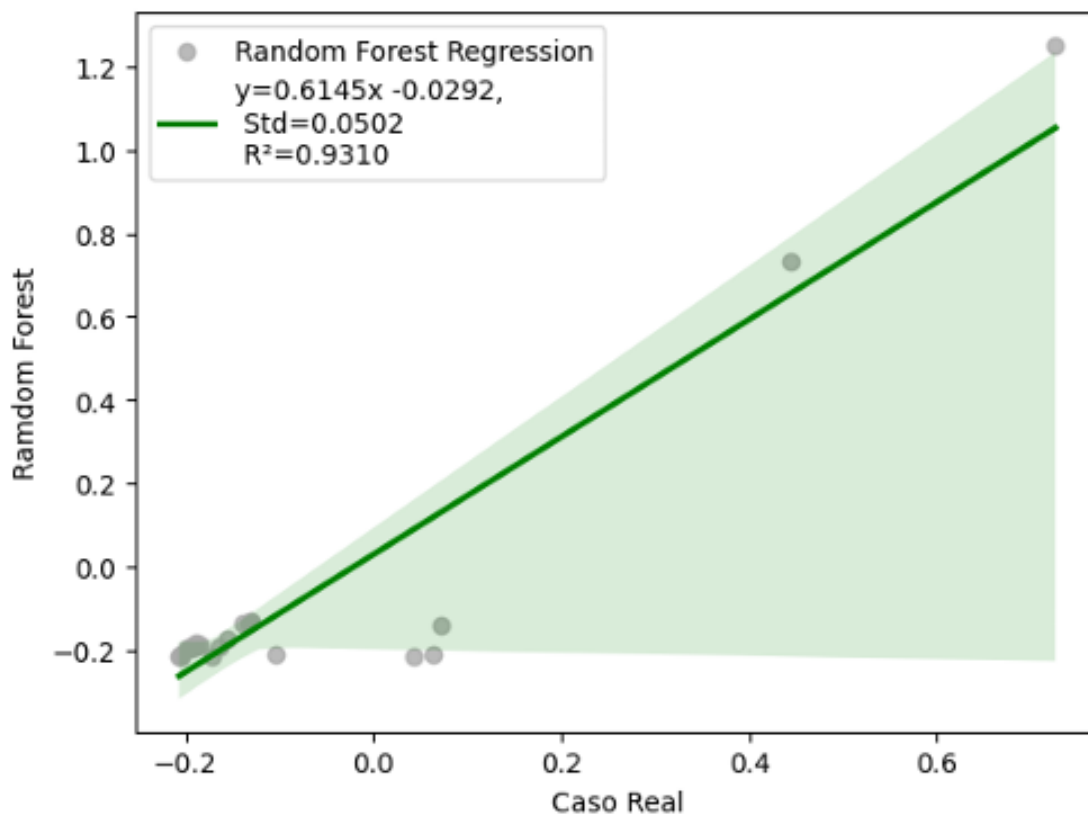


Nota. Regresión lineal entre los datos reales y los resultados de la aplicación del modelo de Regresión Lineal.

En la figura 32, se muestra la regresión lineal de los modelos predictivo de Random Forest o también llamado Bosque Aleatorio. con un ajuste de $R^2 = 0,9310$, quiere decir que predice con un porcentaje igual a 93,10%, en la predicción de coliformes totales mediante este modelo.

Figura 32

Resultado del Modelo de Random Forest



Nota. Regresión lineal entre los datos reales y los resultados de la aplicación del modelo de Random Forest.

5.2 Resultados para los objetivos específicos

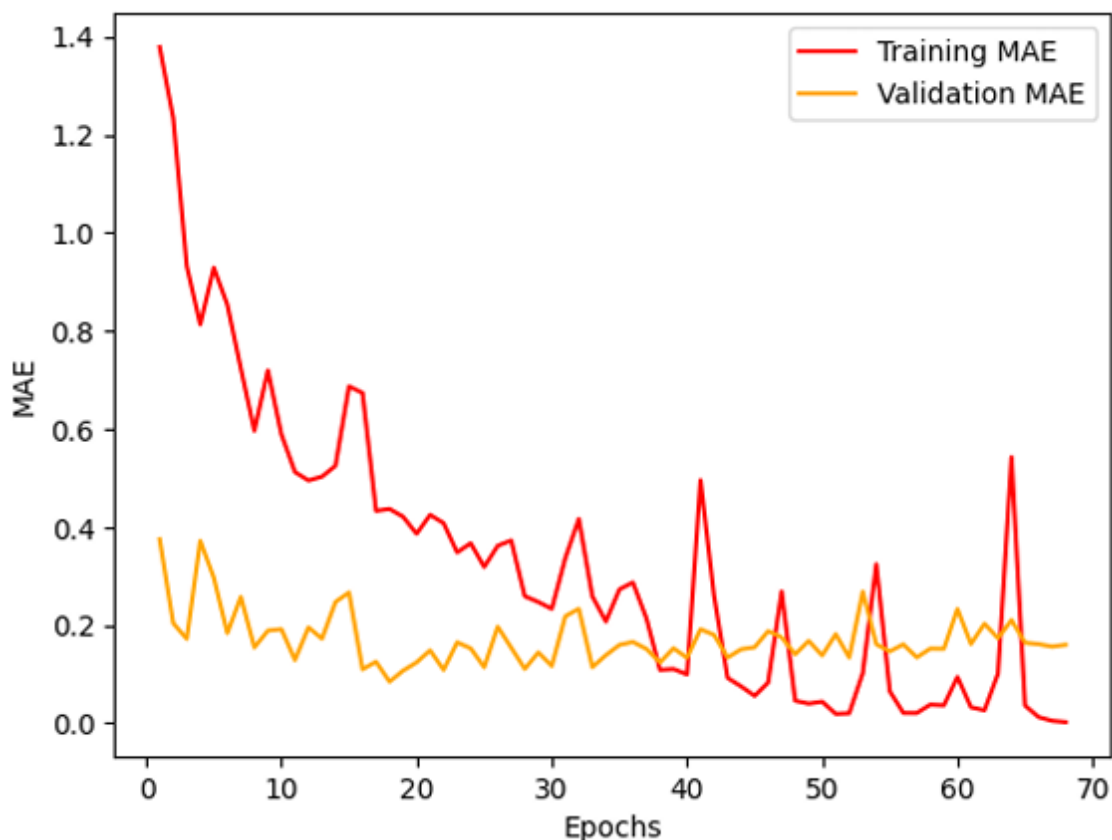
5.2.1 O.E.1: Resultados del objetivo específico 01

Evaluando entre los valores reales de los parámetros microbiológicos y parasitológicos de la calidad del agua y los valores predichos por la aplicación de la inteligencia artificial se obtuvo un grado de correlación $R^2 > 50\%$ en la determinación de los agentes patológicos en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

Se aplicó Mean Absolute Error (MAE) para ver el aprendizaje automático y estadística del modelo, MAE es una métrica de error utilizada para evaluar la precisión de un modelo. Se calcula como el promedio de los valores absolutos de las diferencias entre las predicciones y los valores reales. Es útil para medir el rendimiento de modelos de la Red Neuronal Artificial. En la Figura 33, se muestra los resultados de los errores del aprendizaje del modelo de red neuronal en la predicción de coliformes totales se muestra que en 100 épocas el modelo se redujo de 1.40 hasta el intervalo de [0-0.20] de error, quiere decir que el modelo mientras más épocas se entrena aprende de mejor manera.

Figura 33

Mean Absolute Error (MAE) - Red Neuronal Artificial



Nota. Resultados de la aplicación del modelo de Redes Neuronales Artificiales, métrica MAE.

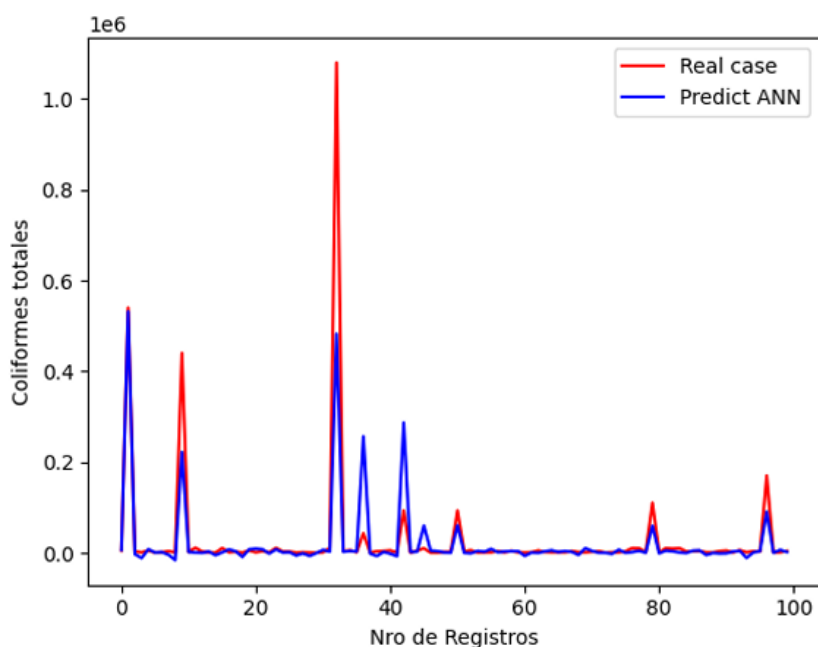
5.2.2 O.E.2: Resultados del objetivo específico 02

Realizando pruebas de entrenamiento y testeo se determinó que el modelo de inteligencia artificial Neuronal Network presenta mayor precisión en la predicción de los parámetros microbiológicos y parasitológicos con respecto al Machine Learning Regression en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024. En la figura 34, se muestra la predicción de coliformes totales mediante el modelo de redes neuronales artificiales en 100 épocas de aprendizaje lo cual mejora el modelo mediante el incremento de épocas de aprendizaje a diferencia de los modelos de aprendizaje supervisado

es que los modelos de redes neuronales se puede ajustar las épocas de aprendizaje y los otros modelos no solo por única vez estos aprenden se puede mejorar el modelo de regresión lineal y Random Forest mas no instruirlo en épocas.

Figura 34

Resultado del Modelo de Redes Neuronales Artificiales



Nota. Resultados de la aplicación del modelo de Redes Neuronales Artificiales en la predicción de coliformes totales.

5.2.3 O.E.3: Resultados del objetivo específico 03

Para realizar las pruebas de entrenamiento y testeo fueron indispensables los parámetros físicos químicos del agua los cuales se aplicaron como datos de entrada en los modelos de inteligencia artificial para determinar la predicción de los parámetros microbiológicos y parasitológicos en la cuenca hidrográfica del río Chumbao, provincia de Andahuaylas, región Apurímac – 2024.

5.3 Discusión de Resultados

La presente investigación tuvo como objetivo principal predecir los agentes patológicos mediante parámetros de calidad del agua aplicando técnicas de Inteligencia Artificial (IA). Tras la recolección y análisis de datos, los resultados demostraron la eficacia de los modelos de IA en la predicción de agentes patológicos a partir de parámetros fisicoquímicos del agua. Entre los modelos utilizados árbol de decisión, bosque aleatorio y redes neuronales artificiales, el algoritmo Random Forest obtuvo el mejor desempeño, con un coeficiente de determinación de $R^2 = 0.9310$, lo que representa una capacidad predictiva del 93.10 % en la estimación de coliformes totales.

Estos resultados coinciden con estudios previos de (Narváez & Tipán, 2022) y (Salazar, 2020) y (Mamani, 2012), quienes reportaron una relación positiva entre el análisis de los parámetros de calidad del agua y la presencia de contaminantes microbiológicos. Según (Narváez & Tipán, 2022), la calidad del agua presenta un índice de 82 en el Punto 1, lo que indica una calidad buena. Los parámetros que sobrepasaron los límites máximos permisibles (LMP) fueron la DBO (> 5 mg/L) y el oxígeno disuelto (> 4 mg/L). Asimismo, el Punto 2 (P2) presenta un índice de 86, lo que también indica una buena calidad de agua, aunque los parámetros DBO (> 5 mg/L), oxígeno disuelto (> 5 mg/L), conductividad eléctrica (> 750 μ S/cm), sólidos disueltos totales (> 500 mg/L) y arsénico (> 0.1 mg/L) superan los LMP. El incremento de actividades sin responsabilidad ambiental y el uso indiscriminado de polímeros han intensificado la contaminación de cuerpos de agua. En este sentido, el presente estudio se justifica por la necesidad de evaluar la calidad del agua del río Pumacunchi y determinar la presencia de microplásticos, con el fin de identificar riesgos ambientales y sanitarios.

Asimismo, los hallazgos obtenidos refuerzan lo propuesto por (Chura, 2020) y (Calo & Guerrero 2019), al confirmar que los componentes de la gestión de la calidad del agua, conforme a los marcos teóricos y normativos, se relacionan estrechamente con la preservación del recurso y la reducción de la contaminación en los ecosistemas acuáticos. Finalmente, se recomienda contar con datos de laboratorio confiables y homogenizados, así como con un conocimiento adecuado de los parámetros de la calidad del agua y de las técnicas de Inteligencia Artificial, para garantizar resultados precisos y coherentes en los análisis, tal como lo sugieren (Albert y Lilia, 1997).

El objetivo específico n°1, de esta investigación fue establecer la correlación entre los parámetros de calidad del agua y la determinación de agentes patológicos mediante laboratorio e Inteligencia Artificial (IA). Los resultados mostraron que la aplicación de estas herramientas permitió obtener un grado de correlación superior al 50 % en la predicción de agentes patológicos. Para evaluar la precisión de los modelos de aprendizaje automático, se utilizó el Mean Absolute Error (MAE), el cual demostró que la Red Neuronal Artificial redujo el error de 1.40 a un rango de 0–0.20 tras 100 épocas de entrenamiento, indicando un aprendizaje progresivo y eficiente del modelo en la predicción de coliformes totales.

Estos hallazgos coinciden con (Vargas, 2020), quien reportó que, tras monitorear diez puntos del río Shilcayo, los parámetros pH (8.1–8.6), temperatura (25.0–26.5 °C) y sólidos suspendidos totales (41 mg/L) se encontraban dentro de rangos aceptables, mientras que la demanda bioquímica de oxígeno (DBO: 75.6 mg/L), demanda química de oxígeno (DQO: 104.9 mg/L) y coliformes termotolerantes (86,890,137.8 NMP/100 mL) superaban los límites máximos permisibles, evidenciando deficiencias en la calidad microbiológica y orgánica del agua.

Asimismo, los resultados respaldan lo indicado por (Rivera & Gómez 2019), quienes afirmaron que la gestión de la calidad del agua basada en marcos teóricos y normativos contribuye significativamente a controlar la presencia de coliformes totales y otros agentes patológicos. Finalmente, se recomienda mantener un registro continuo y actualizado de los parámetros fisicoquímicos y microbiológicos, así como aplicar técnicas de Inteligencia Artificial validadas, para optimizar la capacidad predictiva de los modelos y fortalecer la toma de decisiones en la gestión ambiental (Albert y Lilia, 1997).

El objetivo específico 2, de esta investigación fue aplicar la técnica más eficaz de Inteligencia Artificial para predecir agentes patológicos en la calidad del agua. Los resultados mostraron que, mediante pruebas de entrenamiento y testeo de los parámetros de calidad del agua, las Redes Neuronales Artificiales predicen con mayor precisión los agentes patológicos en comparación con los modelos de regresión y Random Forest. La predicción de coliformes totales, ajustando hasta 100 épocas de aprendizaje, evidenció que el modelo mejora con el incremento de épocas, a diferencia de otros modelos supervisados que no permiten instrucción repetida durante el entrenamiento.

Estos hallazgos coinciden con estudios previos de (Vargas, 2020) y (Vargas, 2021) y (Ramos, 2022), quienes demostraron la efectividad en la evaluación de parámetros físicos, químicos y microbiológicos del agua. La investigación se realizó a lo largo del río, desde la cabecera de la cuenca (13°46'42.5" S, 73°13'50.0" O, 4 295 m) hasta el último punto de muestreo (13°35'26.4" S, 73°27'0.8" O, 2 529 m), considerando pendientes medias de 4%, evaluando indicadores como oxígeno disuelto, DBO5, nitratos, fosfatos, turbidez, temperatura, pH, conductividad y coliformes totales. Los análisis estadísticos incluyeron ANOVA, test de

Tukey y correlación de Pearson ($p < 0,05$), identificando zonas con calidad excelente, aceptable y excesivamente contaminada.

Los hallazgos respaldan lo planteado por (Palau, 2014), al evidenciar que los parámetros de calidad del agua son fundamentales para la predicción de agentes patológicos y coliformes totales mediante Redes Neuronales Artificiales. Esto confirma que la gestión de la calidad del agua, basada en marcos teóricos y normativos, contribuye a la preservación del recurso y a la disminución de la contaminación en los ecosistemas acuáticos. Se recomienda mantener un monitoreo continuo y actualizado de los parámetros fisicoquímicos y microbiológicos, junto con la aplicación de modelos de Inteligencia Artificial validados, para optimizar la predicción de agentes patológicos y la toma de decisiones en la gestión ambiental (Albert y Lilia, 1997).

VI. Conclusiones

- ✓ La implementación de modelos de Inteligencia Artificial en la cuenca del río Chumbao demuestra que la integración de parámetros fisicoquímicos permite una predicción altamente precisa de agentes patológicos con una seguridad del 95% podemos afirmar que los datos se encuentran muy similares en un 98.72% entre sus valores. En la Figura 29, nos muestra una regresión lineal que hace la comparación de los datos reales y el modelo de red neuronales artificial y el resultado de $R^2 = 98,87\%$ esto nos indica que el modelo de red neuronal artificial predice valores de coliformes totales con una seguridad del 95% podemos afirmar que los datos se encuentran muy similares en un 98.87% entre sus valores. Esta capacidad predictiva se establece como una herramienta crítica para la gestión del riesgo sanitario en la provincia de Andahuaylas, permitiendo una respuesta proactiva ante las variaciones de la calidad del agua.
- ✓ Se aplicó Mean Absolute Error (MAE) para ver el aprendizaje automático y estadística del modelo, MAE es una métrica de error utilizada para evaluar la precisión de un modelo. Se calcula como el promedio de los valores absolutos de las diferencias entre las predicciones y los valores reales. Es útil para medir el rendimiento de modelos de la Red Neuronal Artificial. En la Figura 34, se muestra los resultados de los errores del aprendizaje del modelo de red neuronal en la predicción de coliformes totales se muestra que en 100 épocas el modelo se redujo de 1.40 hasta el intervalo de [0-0.20] de error, quiere decir que el modelo mientras más épocas se entrena aprende de mejor manera. Por ello, los resultados obtenidos demuestran una correspondencia significativa entre los registros de campo y las

estimaciones del modelo de inteligencia artificial, alcanzando un coeficiente de determinación R^2 muy similares en un 98.72%. Este valor valida la eficacia de las redes neuronales para capturar la variabilidad de los agentes microbiológicos y parasitológicos en la cuenca del río Chumbao, consolidando al modelo como una herramienta de monitoreo predictivo fiable para la región Apurímac.

- ✓ Tras exhaustivas fases de entrenamiento y validación, se concluye que las Redes Neuronales Artificiales (ANN) exhiben una precisión predictiva superior frente a los modelos de Regresión de Machine Learning convencional. Esta ventaja radica en la capacidad de las ANN para modelar las interacciones no lineales complejas de los parámetros microbiológicos y parasitológicos en la cuenca del río Chumbao, superando las limitaciones de los métodos lineales y consolidándose como la arquitectura más robusta para la vigilancia sanitaria en Andahuaylas.
- ✓ Se concluye que la integración de los parámetros fisicoquímicos como variables de entrada (input) fue determinante para el éxito de los modelos de inteligencia artificial. Esta sinergia permitió que los algoritmos decodificaran la relación multidimensional entre el entorno físico del agua y la proliferación de agentes microbiológicos y parasitológicos, estableciendo una metodología de predicción indirecta altamente eficiente para la gestión hídrica de la cuenca del río Chumbao.

VII. Recomendaciones

- ✓ Se recomienda la creación de una plataforma digital que utilice el modelo de Redes Neuronales Artificiales para emitir alertas en tiempo real. De tal manera la Red de Salud de Andahuaylas y a las autoridades competentes de la cuenca del río Chumbao, adoptarían como una herramienta complementaria a la vigilancia epidemiológica. La capacidad predictiva demostrada permite anticipar riesgos sanitarios asociados a parámetros microbiológicos y parasitológicos, facilitando la toma de decisiones proactivas para la prevención de enfermedades de transmisión hídrica en la región Apurímac.
- ✓ Dada la alta correlación entre los parámetros fisicoquímicos y la carga patógena detectada, se sugiere priorizar el equipamiento de estaciones de monitoreo automatizadas en puntos estratégicos de la cuenca. Esto permitiría la obtención de datos fisicoquímicos en tiempo real que, al ser procesados por el modelo de inteligencia artificial, reducirían significativamente los costos operativos y el tiempo de respuesta frente a eventos de contaminación microbiológica, optimizando así el presupuesto destinado a análisis de laboratorio complejos.
- ✓ Para futuras investigaciones, se recomienda expandir el conjunto de datos (dataset) incorporando mediciones multianuales que abarquen con mayor detalle las transiciones entre las épocas de estiaje y lluvias en la sierra peruana. Se estima que la integración de variables climáticas adicionales, como la intensidad de precipitación y la tasa de escorrentía, podría incrementar el coeficiente de determinación (R^2) del modelo actual.

- ✓ Se sugiere realizar un estudio para identificar cuál de los parámetros fisicoquímicos (pH, turbiedad, conductividad eléctrica, DBO, dureza total, nitratos, temperatura) tiene mayor peso en la predicción. Esto permitiría priorizar la compra de sensores específicos para esos parámetros en la cuenca.
- ✓ Se sugiere evaluar la aplicabilidad de los modelos entrenados en la cuenca del río Chumbao hacia otras microcuencas de la región Apurímac con características geomorfológicas similares. Esta práctica de transfer learning (aprendizaje por transferencia) permitiría validar la capacidad de generalización del modelo y fomentaría la creación de un sistema regional unificado de gestión de recursos hídricos basado en inteligencia artificial.
- ✓ Es fundamental recomendar programas de capacitación técnica dirigidos a los gestores de recursos hídricos y técnicos municipales de Andahuaylas en el manejo de herramientas de aprendizaje automático. El éxito de la implementación de modelos de redes neuronales depende no solo de la precisión algorítmica, sino de la capacidad de los operadores locales para interpretar y aplicar los resultados en los planes de ordenamiento territorial y protección de fuentes de agua.
- ✓ Nuestra investigación aporta un método para inferir parámetros microbiológicos (caros y lentos) mediante parámetros fisicoquímicos (baratos y rápidos), lo que democratiza el monitoreo de la calidad del agua en zonas de bajos recursos.
- ✓ Se recomienda implementar técnicas de vanguardia en inteligencia artificial a problemas rurales específicos de la sierra peruana, rompiendo la brecha tecnológica en la gestión de recursos hídricos.

VIII. Referencias

Agua Rio. (3 de julio de 2017). *Fondo para la Comunicación y la Educación Ambiental*.

Obtenido de Agua (Centro de Información del Agua):
<https://agua.org.mx/propiedades-derl-agua>

Aguirre y Diaz. (2018). Efectos de fármacos humanos en *Corbicula fluminea* . Evaluación de neurotoxicidad, inflamación, actividad gametogénica y estado energético. *Ecotoxicología y Seguridad Ambiental*, 652 - 663. Obtenido de <https://cdnsiencepub.com/doi/epdf/10.1139/er-2020-0054>

Albert y Lilia. (1997). *Nitratos y nitritos. Introducción a la Toxicología ambiental*”Centro panamericano de ecología humana y salud. División de salud. Ciudad de México: Lilia A. Albert. Obtenido de <https://es.scribd.com/document/379791389/Introduccion-a-la-Toxicologia-Ambiental>

Apaza, Gárate, Cuadros y Ccasa. (2022). Predictive model of water potability through a decision tree in Artificial Intelligence. En S. S. Ángel Alexis Zevallos Apaza, *Innovación y Sofwae* (págs. 121–131). Arequipa: La salle. Obtenido de <https://revistas.ulasalle.edu.pe/innosoft/article/view/72>

Arias y Holgado. (2022). *Metodología de la investigación: El método ARIAS para desarrollar un proyecto de tesis*. Puno: Creative Commons Atribución 4.0. Obtenido de <https://editorial.inudi.edu.pe/index.php/editorialinudi/catalog/book/22>

Arrojo, A. P. (15 de diciembre de 2022). *Declaración de final de misión del Relator Especial para los derechos humanos al agua potable y saneamiento, señor Pedro Arrojo Agudo*. Obtenido de

<https://www.ohchr.org/sites/default/files/documents/issues/water/statements/2022-12-14/2022-12-15-eom-statement-peru-sr-water-sanitation-sp.pdf>

Arteaga y García. (2019). Hepatitis A: Epidemiología y estrategias de intervención. *Medicina legal de Costa Rica*, 7. Obtenido de <https://www.scielo.sa.cr/pdf/mlcr/v36n2/2215-5287-mlcr-36-02-101.pdf>

Asociación Española de Abastecimientos de Agua y Saneamiento, A. (2020). COVID-19: Agua y saneamiento. <https://www.esamur.com/public/file/COVID19AEAS3abril2020.pdf>.

Baquerizo y Acuña. (2019). CONTAMINACIÓN DE LOS RIOS: CASO RIO GUAYAS Y SUS AFLUENTES. *Revista de Investigación Científica Universidad Nacional de Tumbes*, 8. Obtenido de <https://revistas.untumbes.edu.pe/index.php/manglar/article/view/118>

Barcelo, D. (2003). "Emerging pollutants in water analysis". *Digital Universitaria*, 18 - 26. Obtenido de <http://193.122.196.39:8080/xmlui/handle/10521/2559>

Becerril. (2009). Contaminantes emergentes en el agua. *Digital Universitaria*, 3 - 50. Obtenido de <https://www.revista.unam.mx/vol.10/num8/art54/int54.htm>

Bengio y Hinton. (10 de Septiembre de 2015). *The Quartz guide to artificial intelligence: Deep learning*. Nueva York: IGI Global Scientific Publishing. Obtenido de <https://qz.com/1046350/the-quartz-guide-to-artificial-intelligence-what-is-it-why-is-it-important-and-should-we-be-afraid>

Bernaola y Chávez. (2011). Guía de práctica clínica sobre el diagnóstico y tratamiento de la diarrea aguda infecciosa en pediatría, Perú - 2011. *Revista de Gastroenterología del Perú*, 258 - 277. Obtenido de <http://www.scielo.org.pe/pdf/rgp/v31n3/a09v31n3.pdf>

- Bernaola, G., & Luque, W. (2002). Fisiopatología de las infecciones por Adenovirus. *Asociación de Médicos Residentes del Instituto de salud del Niño*, 41 - 47. Obtenido de https://sisbib.unmsm.edu.pe/bvrevistas/paediatria/v04_n2/pdf/fisiopatologia_adenovirus.pdf
- Bernardina, E. (1986). *"Tecnología de aceites y grasas"*. Madrid: Alhambra Madrid. Obtenido de Acriba S.A.: <https://repositorio.unan.edu.ni/3504/1/61591.pdf>
- Bhatti y Latif. (2011). Evaluación de la calidad del agua de un río mediante un enfoque de indexación durante la temporada de bajo caudal. *Irrigation and Drainage*, 11. Obtenido de <https://www.sciencedirect.com/science/article/pii/S2665972724000850>
- Blasco. (1998). Análisis de Aguas y Desagües. *Lima Uni*, 248. Obtenido de https://biblioteca.unasam.edu.pe/cgi-bin/koha/opac-detail.pl?biblionumber=18662&query_desc=an%3A21978
- Blasco, E. J. (1998). *"Análisis de aguas y desagües"*. Lima: Banco de Libros. Obtenido de <https://biblioteca.unasam.edu.pe/bib/18662>
- Blume y Macedo. (2010). Evaluación de la calidad del agua del río Sinos, sur de Brasil. *Brazilian Journal Of Biology*, 8. Obtenido de https://www.researchgate.net/publication/49745979_Water_quality_assessment_of_the_Sinos_River_Southern_Brazil
- Boston. (05 de febrero de 2019). *E. Coli (Escherichia Coli)*. Obtenido de City of Boston: <https://www.boston.gov/government/cabinets/boston-public-health-commission/infectious-diseases/infectious-diseases-z/shiga-toxin-producing-e-coli-stec>

Cabezas, C. (2018). Enfermedades infecciosas relacionadas con el agua en el Perú. *Revista Peruana de Medicina Experimental y Salud Pública*. Perú: <https://doi.org/10.17843/rpmesp.2018.352.3761>.

Calidad Ambiental. (07 de junio de 2017). Aprueban los Estándares de Calidad Ambiental para Agua y Disposiciones Complementarias - Decreto supremo N° 004. *El Peruano*, pág. 10.

Calidad Ambiental. (7 de Junio de 2017). *Calidad Ambiental para agua y establecen disposiciones complementarias (Decreto Supremo N 004-2017-MINAM)*. Obtenido de El Peruano: <https://www.minam.gob.pe/wp-content/uploads/2017/06/DS-004-2017-MINAM.pdf>

Campbell y Reece. (2007). *Biología Ed. Médica Panamericana*. Estados Unidos: Médica Panamericana.

Campbell y Stanley. (2015). *Experimental and quasi-experimental designs for research*. Ravenio books. Chicago: Libros Ravenio. Obtenido de https://books.google.com.pe/books?hl=es&lr=&id=KCTrCgAAQBAJ&oi=fnd&pg=PT12&dq=Experimental+and+quasi-experimental+designs+for+research.+Ravenio+books.+Chicago:+Libros+Ravenio.&ots=Mfq4GTqBSI&sig=nYE5BZULxi3j_Am---mPPSXEUMw#v=onepage&q&f=false

Campbell, N. A., & Reece, J. B. (2007). *Biología. Ed. Médica Panamericana*, 16-18. Obtenido de <https://books.google.com.co/books?id=QcU0yde9PtkC&printsec=copyright#v=onepage&q&f=false>

Campos, M., Beltrán, M., Fuentes, N., & Moreno, G. (2018). Huevos de Helmitos como indicadores de contaminación de origen fecal en aguas de riego agrícola, biosólidos, suelos y pastos. *Biomédica*. <https://doi.org/10.7705/biomedica.v38i0.3352>.

Carcelèn y Moran. (26 de Abril de 2022). El impacto de la epidemia de cólera de 1991 en la costa norte del Perú. *Creative Commons*, 1 - 14. Obtenido de http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2415-09592022000100003

Carrillo, & Hilaño, P. A. (05 de Septiembre de 2023). "Evaluación de la calidad de agua a través de macro invertebrados bentónicos en la cuenca del río Kushapukos en el cantón Tiwintza, provincia de Morona Santiago". *La Libertad: Universidad Estatal Península de Santa Elena*, 2023, 58. Obtenido de <https://repositorio.upse.edu.ec/handle/46000/10122>

Carta del Agua. (06 de Mayo de 1968). "Carta Europea del Agua". *Canal Educa*, pág. 1. Obtenido de <https://www.fundacioncanal.com/canaleduca/wp-content/uploads/2015/09/Anexo-III-Carta-Europea-del-Agua.pdf>

Clark. (2017). The Journal of Positive Psychology. En V. L. Plano Clark, *Journal of Positive Psychology, Volume 12, Number 3* (págs. 305 - 306). Detroit: La Revista de Psicología Positiva. Obtenido de <https://www.tandfonline.com/toc/rpos20/12/3#:~:text=0Altmetric-Full%20Access,Danielle%20Berado%20&%20Terese%20Jean%20Lund>

Cohen y Aiken. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Routledge. Obtenido de <https://www.taylorfrancis.com/books/mono/10.4324/9780203774441/applied->

multiple-regression-correlation-analysis-behavioral-sciences-jacob-cohen-patricia-cohen-stephen-west-leona-aiken

Comité Asesor de Vacunas, C. (2018). *Fiebre Tifoidea. Manual de vacunas en línea de la AEP.* Asociación Española de Pediatría: <https://vacunasaep.org/documentos/manual/cap-25#8>.

Conagua. (11 de Noviembre y Diciembre de 2013). Ceremonia de Clausura del Año Internacional de la Cooperación en la ESFERA DEL AGUA. *Agenda Internacional Noviembre - Diciembre, 2.* Obtenido de https://www.gob.mx/cms/uploads/attachment/file/134817/Agenda_Internacional_Nov_Dic.pdf

Conagua, C. N. (2018). *Red Nacional de Monitoreo de la Calidad de las Aguas Nacionales de Conagua.* Obtenido de <https://datos.gob.mx/busca/dataset/red-nacional-de-monitoreo-de-la-calidad-de-las-aguas-nacionales-de-conagua>

Condori. (15 de Setiembre de 2020). *Universo, población y muestra.* Obtenido de Acta de Academia: <https://www.aacademica.org/cporfirio/18.pdf>

Cook y Donald. (2007). *Experimental and quasi-experimental designs for generalized causal inference.* San Francisco: Houghton Mifflin Compan. Obtenido de <https://iaes.cgiar.org/sites/default/files/pdf/147.pdf>

Crespo y Rivero. (08 de Agosto de 2008). Hepatitis E. *Revista Española de ENFERMEDADES DIGESTIVAS,* 514. Obtenido de <https://scielo.isciii.es/pdf/diges/v100n8/paciente.pdf>

Creswell y Creswell. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches.* Los angeles: SAGE. Obtenido de

[https://books.google.com.pe/books?hl=es&lr=&id=335ZDwAAQBAJ&oi=fnd&pg=PT16&dq=Creswell+y+Creswell.+\(2017\).+Research+design:+Qualitative,+quantitative,+and+mixed+methods+approaches.+Los+angeles:+SAGE.&ots=YExTGItsuF&sig=c5_vxEa4aJ0utP7E22N7idayckQ#v=onepag](https://books.google.com.pe/books?hl=es&lr=&id=335ZDwAAQBAJ&oi=fnd&pg=PT16&dq=Creswell+y+Creswell.+(2017).+Research+design:+Qualitative,+quantitative,+and+mixed+methods+approaches.+Los+angeles:+SAGE.&ots=YExTGItsuF&sig=c5_vxEa4aJ0utP7E22N7idayckQ#v=onepag)

Cuya y García. (2022). "Evaluación de la calidad ambiental del agua potable y percepción en la comunidad de Huaychao, Acos Vinchos - Huamanga - Ayacucho, 2022". *Universidad Cesar Vallejo*, 1-71. Obtenido de <https://core.ac.uk/download/548503814.pdf>

De la Fuente, C. (07 de Noviembre de 2013). *Parámetros físicoquímicos del agua*. Madrid: Grupo Asis. Obtenido de Dialnet: https://www.adiveter.com/ftp_public/A3081113.pdf

Díaz Castrillón, F., & Toro Montoya, A. (2020). SARS-CoV-2/COVID-19: El virus, la enfermedad y la pandemia. *Medicina y Laboratorio*. <https://doi.org/10.36384/01232576.268>.

Diersing, N. (2009). "Water quality: Frequently asked questions". *Florida Brooks National Marine Sanctuary, Key West, FL*, 1-2. Obtenido de <https://nmsfloridakeys.blob.core.windows.net/floridakeys-prod/media/archive/scisummaries/wqfaq.pdf>

Disachanka. (2022). *Análisis Situacional de Salud, 2022*. Obtenido de Disachanka: https://disachanka.gob.pe/wp-content/uploads/2024/02/ASIS-2022-DISTRITO-ANDAHUAYLAS-OK_FIN.pdf

Don y Perry. (2008). *Manual del ingeniero químico de Perry, octava edición*. New York, US: Access Engineering. Obtenido de <https://www.accessengineeringlibrary.com/content/book/9780071422949>

- Espigares. (2006). Virus en aguas de consumo. *Campus Universitario de Cartuja*. 18071 Granada, 173 - 189. Obtenido de [https://saludpublica.ugr.es/sites/dpto/spublica/public/inline-files/bc510159f5f2fa3_Hig.Sanid_.Ambient.6.173-189\(2006\).pdf](https://saludpublica.ugr.es/sites/dpto/spublica/public/inline-files/bc510159f5f2fa3_Hig.Sanid_.Ambient.6.173-189(2006).pdf)
- Espinoza, K. A. (Octubre de 2020). "Evaluación de la calidad de agua y metales traza del río Jubones en el Cantón Pasaje, provincia de El Oro, Ecuador". *Facultad de Ciencias Naturales. Universidad de Guayaquil*, 21 -71. Obtenido de <https://repositorio.ug.edu.ec/bitstreams/61343de9-dc05-4143-9753-5d8d02729b40/download>
- Finlay. (2014). *Análisis Predictivo, Minería de Datos y Big Data. Mitos, conceptos erróneos y métodos (1ª ed.)*. Chennai: Palgrave Macmillan. Obtenido de Palgrave Macmillan: https://www.researchgate.net/publication/263617471_Predictive_Analytics_Data_Mining_and_Big_Data_Myths_Misconceptions_and_Methods
- Froehner y Da Rosa. (7 de Octubre de 2010). Seguimiento de los aportes antropogénicos en el río Barigüi, Brasil, mediante biomarcadores. *Water, Air, & Soil Pollution*, 210 , 33 - 41, 8. Obtenido de <https://aidisnet.org/wp-content/uploads/2019/08/02-salud-y-medio-ambiente.pdf>
- García y Arias. (11 de Octubre de 2004). Virus en sistemas acuáticos e implicaciones en la Salud Pública. *Hidrobiológica vol.14 no.2* , 16-18. Obtenido de https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-88972004000200011
- Guerra. (2001). *Aspectos Sanitarios del Uso de Aguas Residuales. CEPIS. Organización Panamericana de la Salud (OPS)*, 1-10. Rio de Janeiro: Organización Panamericana de la Salud (OPS). Obtenido de

https://iris.who.int/bitstream/handle/10665/39333/who_tr_s_778_spa.pdf;jsessionid=7EFE95AC1E3A0AF463FD7596FA4EBC63?sequence=1

Harvez y Àvila. (2013). *La epidemia de C3lera en Am3rica Latina: reemergencia y morbilidad*.

Revista Panamericana de Salud P3blica: <https://doi.org/10.1590/S1020-49892013000100006>.

Hernàndez y Fernand3z. (2014). *Metodolog3a de la investigaci3n (Vol. 6)*. Ciudad de M3xico:

McGRAW-HILL. Obtenido de <https://www.esup.edu.pe/wp-content/uploads/2020/12/2.%20Hernandez,%20Fernandez%20y%20Baptista-metodolog%C3%ADa%20Investigacion%20Cientifica%206ta%20ed.pdf>

Instituto Nacional de Salud, I. (2019). Hepatitis A. *Bolet3n epidemiol3gico semanal*.

<https://www.ins.gov.co/Noticias/Dengue/5.%202019%20Bolet%C3%ADn%20epidemio%C3%B3gico%20semana%207.pdf>.

Johnson y Bassett. (1997). "Meanings of environmental terms". *Journal of Environmental*

Quality, 144 -148. Obtenido de https://www.researchgate.net/publication/240784159_Meanings_of_Environmental_Terms

Khan y Shamim. (2022). *Comprensi3n del mecanismo de resistencia a los antimicrobianos y*

patog3nesis de Salmonella enterica serovar typhi. Lahore: <https://doi.org/10.3390/microorganisms10102006>.

Levin. (2006). *Study design iii: Cross-sectional studies*. New York: Evid Based Dent 7.

Obtenido de <https://www.nature.com/articles/6400375>

Lopez. (20 de Diciembre de 2018). "Agentes patol3gicos". Obtenido de Coliformes:

caracter3sticas, tipos, g3neros, enfermedades: <https://www.lifeder.com/coliformes/>

- Maiquiza, K. E., & Tonato, G. K. (Septiembre de 2020). "Identificación de Diatomeas Epilíticas asociadas a la calidad del agua en el río Yanayacu, sector San Juan, cantón Salcedo, provincia de Cotopaxi, 2020". *Ecuador, Latacunga: Universidad Técnica de Cotopaxi UTC.*, 10-11. Obtenido de <https://repositorio.utc.edu.ec/server/api/core/bitstreams/941735de-3ffd-4e70-9d66-bf51d981f013/content>
- Mamani. (Julio de 2017). "Calidad del agua del río Cumbaza respecto al efluente de la planta de tratamiento, San Roque de Cumbaza, 2021". *UNAP-Institucional*, 21-20. Obtenido de <http://repositorio.unap.edu.pe/handle/20.500.14082/6747>
- Mamani, L. (2012). Presencia de protozoarios y helmitos en agua de consumo humano de la región Moquegua. Tacna, Perú: http://repositorio.unjbg.edu.pe/bitstream/handle/UNJBG/1940/104_2012_mamani_mamani_lv_faci_biologia_microbiologia.pdf?sequence=1&isAllowed=y.
- Marin y Marin. (2022). "Impacto de las fuentes de contaminación ambiental en la calidad de agua del río Llaucano, Bambamarca, Cajamarca, 2021". *UCV-Institucional*. Obtenido de https://alicia.concytec.gob.pe/vufind/Record/UCVV_c33274b86ebe8e1be51d094f948d7ecf
- Menard. (2002). *Longitudinal Research (Quantitative Applications in the Social Sciences)*. San Francisco: SAGE Publications. Obtenido de https://link.springer.com/chapter/10.1007/978-3-319-08720-7_7
- MINAM. (2017). *Decreto Supremo N° 004-2017-MINAM*. Perú: Minam. Obtenido de <https://www.minam.gob.pe/disposiciones/decreto-supremo-n-004-2017-minam/>

Minan. (30 de Julio de 2008). Aprueban los estándares nacionales de calidad ambiental para agua, DS. N° 002. *El Peruano*, pág. 12. Obtenido de https://www.ana.gob.pe/sites/default/files/normatividad/files/ds_002_2008_minam_0_0_2.pdf

Minaya y Rodríguez. (2018). Evaluación Físico-Química y Microbiológica de Agua para consumo humano Puno - Perú. *Fides et Ratio - Revista de Difusión cultural y científica de la Universidad La Salle en Bolivia*, 47 - 68. Obtenido de http://www.scielo.org.bo/scielo.php?pid=S2071-081X2018000100005&script=sci_arttext

Ministerio de Salud. (2023). *Plan de abordaje integral de la enfermedad diarreica aguda y plan de contingencia de Cólera*. Buenos Aires: Ministerio de Salud - Presidencia de la Nación. Obtenido de <https://www.infobioquimica.com/multimedia/noticias/guia-abordaje-colera.pdf>

Ministerio de Salud, M. (2013). Esquema Nacional de Salud. http://www.minsa.gob.pa/sites/default/files/programas/esquema_de_vacunacion_revisado_marzo_2013.pdf.

Minsa. (26 de Setiembre de 2010). *Reglamento de la calidad del agua para consumo humano: Ds N° 031-2010-sa*. Obtenido de Biblioteca Nacional del Perú: <https://cdn.www.gob.pe/uploads/document/file/273650/reglamento-de-la-calidad-del-agua-para-consumo-humano.pdf?v=1561937448>

Minsa. (2018). *Situación epidemiológica de las enfermedades diarreicas agudas en el Perú*. Boletín Epidemiológico del Perú. Obtenido de <https://www.dge.gob.pe/portal/docs/vigilancia/boletines/2018/19.pdf>

- Miranda. (2013). El Cólera (una amenaza latente en tiempos de desastres). *Rev Médica panacea*, 25 - 26. Obtenido de <https://revistas.unica.edu.pe/index.php/panacea/article/view/126/120>
- Mohammeed, Hameed y Seidu. (2017). *Random forest tree for predicting fecal indicator organisms in drinking water supply*. Cracovia: IEEE. Obtenido de <https://ieeexplore.ieee.org/abstract/document/8256398>
- Moncada y Luis. (2016). *Satisfacción de los usuarios sobre la atención brindada en el servicio de fisioterapia de consulta externa, del hospital Alemán Nicaraguense*. Managua: Repositorio Institucional UNAM - Managua. Obtenido de <https://core.ac.uk/download/pdf/80118544.pdf>
- Mophin-Kani y Murugesan. (2010). Evaluación y clasificación de la calidad del agua del río perenne Tamirabarani mediante la agregación del índice de calidad del agua. *Revista de Biología Básica y Aplicada*, 209-215. Obtenido de <https://www.scirp.org/reference/referencespapers?referenceid=2709612>
- Mora, Martínez y Figueroa. (2010). Protozoarios en aguas superficiales y muestras fecales de individuos de poblaciones rurales del municipio Montes, estado de Sucre, Venezuela. *Investigación clínica V.51 - N.4 Maracaibo*, 9-12. Obtenido de https://ve.scielo.org/scielo.php?pid=S0535-51332010000400003&script=sci_arttext
- Moscoso. (2016). *Manual de buenas prácticas para el uso seguro y productivo de las aguas residuales domésticas*. A nivel Nacional - Perú: Autoridad Nacional del Agua (ANA). Obtenido de <https://repositorio.ana.gob.pe/handle/20.500.12543/207>
- Muntané. (2010). "Introducción a la Investigación Básica". *Instituto de Salud Carlos III*, 1 - 7. Obtenido de _____ de _____

https://www.researchgate.net/publication/341343398_Introduccion_a_la_Investigacion_basica

Muñoz, C. (2008). Caracterización y tratamiento de aguas residuales. *Instituto de Ciencias Básicas e Ingeniería*, 42 - 44. Obtenido de <http://dgsa.uaeh.edu.mx:8080/bibliotecadigital/bitstream/handle/231104/514/Caracterizacion%20y%20tratamiento%20de%20aguas%20residuales.pdf?sequence=1&isAllowed=y>

Narváez, J. J., & Tipán, S. A. (Agosto de 2022). "Análisis de la calidad de agua y presencia de microplásticos en la parte alta, media y baja del río Pumacunchi en la provincia de Cotopaxi, periodo 2022". Latacunga, Ecuador. Obtenido de <https://repositorio.utc.edu.ec/server/api/core/bitstreams/6eed5271-612e-49b2-bbe7-e2e135fcd7c/content>

Nicomedes. (30 de Agosto de 2018). *Tipos de Investigación*. Obtenido de UNISDG-Institucional: <http://repositorio.usdg.edu.pe/handle/USDG/34>

Norduzian, M. (Septiembre de 1984). *Demanda bioquímica de oxígeno, conceptos y metodología*. Ciudad de México: Instituto de Ingeniería UNAM. Obtenido de <http://www2.iingen.unam.mx/es-mx/difusion/CongresoSMBB/SMBB/Libros/D-22.pdf>

OMS. (1998). Guía para la calidad del agua potable, V3: Vigilancia y control de los abastecimientos de agua a la comunidad de huaychao. *Organización Mundial de la Salud*. Obtenido de <https://iris.who.int/bitstream/handle/10665/41985/9243545035-spa.pdf?sequence=1&isAllowed=y>

- Organización Panamericana de la Salud, O. (2018). Alerta epidemiológica: Salmonella entérica serovar Typhi Haplotipo H58. *Organización Mundial de la Salud*. <https://vacunasaep.org/documentos/manual/cap-25>.
- Perez y Mariscal. (1999). *Coagulación y floculación. Cap. 11 de Estudio Sanitario del Agua*. Granada, España: J.A. y Espigares García M. (eds). Editorial. Obtenido de https://cidta.usal.es/cursos/simulacion/modulos/libros/Pasar/coagulacion_floculacion.pdf
- Pino, H. M. (03 de Agosto de 2017). *Transitioning entirely to neural machine*. Obtenido de <https://code.facebook.com/posts/289921871474277/transitioning-entirely-to-neural-machinetranslation>
- Quiñones, Ochoa y Gamarra. (2020). *Red neuronal artificial para estimar un índice de calidad de agua*. Quito: Universidad UTE. Obtenido de <https://www.redalyc.org/journal/5722/572262509013/html/>
- Radulovich, R. (1997). "Sostenibilidad en el uso del agua en América Latina". *Revista Forestal Centroamericana (CATIE)*, 13 - 17. Obtenido de <http://bco.catie.ac.cr:8087/portal-revistas/index.php/RRNA/article/view/1243>
- Ramirez y Duran. (2008). *Demanda química de oxígeno de muestras acuosas*. Obtenido de Landy Irene Ramírez Burgos: http://cmas.siu.buap.mx/portal_pprd/work/sites/redica/resources/LocalContent/127/2/Libro%20DQO%202008.pdf
- Ramos, B. S. (2022). "Formulación de índices de contaminación del agua (ICO) para un río altoandino: caso subcuenca Chumbao, Andahuaylas, periodo 2018-2019". Cusco,

Perú. Obtenido de <https://repositorio.uandina.edu.pe/item/17711aa3-2291-4b95-89f8-769a6dfbb2d6>

Raymundo y Rafaela. (2018). Incidencia de Fiebre Tifoidea, Fiebre Paratifoidea y Fiebre de Malta en pobladores de AA.HH. - Villa María del Triunfo, año 2018. *Universidad Norbert Wiener*, 14 - 15. Obtenido de <https://repositorio.uwiener.edu.pe/server/api/core/bitstreams/132d602c-8e74-4eef-9566-96cbd631e3f4/content>

Reina, J., & Dueñas, J. (2019). Detección de Parechovirus Humanos en muestras respiratorias en menores de meses con fiebre de origen desconocido. *Revista Española de Quimioterapia*. Obtenido de <https://pdfs.semanticscholar.org/8ac9/69bc99019daa095ffecb1b7f9de4c44e4419.pdf>

Restrepo Gutierrez, J., & Toro Montoya, A. (2011). Hepatitis A. *Medicina & Laboratorio*. Obtenido de <https://www.redalyc.org/articulo.oa?id=716479633002>

Ríos Tobón, S., Agudelo Cadavid, R. M., & Gutierrez Builes, L. A. (2017). Patógenos e indicadores microbiológicos de calidad de agua para consumo humano. Colombia. Obtenido de <http://www.scielo.org.co/pdf/rfnsp/v35n2/0120-386X-rfnsp-35-02-00236.pdf>

Rodríguez Morales, A., Sánchez Duque, J., Hernández Botero, S., Pérez Díaz, C., Villamil Gómez, W., Méndez, C., . . . Paniz Mondolfi, A. (2020). Preparación y control de la enfermedad por Coronavirus 2019 (COVID-19). *América Latina*: <https://doi.org/10.35663/amp.2020.371.909>.

Rouhiainen. (2018). *Inteligencia Artificial*. Madrid: Alienta. Obtenido de Editorial Planeta, S.A.:
https://planetadelibrosec0.cdnstatics.com/libros_contenido_extra/40/39308_Inteligencia_artificial.pdf

Rouhiainen. (2018). *Inteligencia Artificial*, Ed. Alienta 20-21. Madrid: Alienta. Obtenido de
https://proassetspdlcom.cdnstatics2.com/usuarios/libros_contenido/arxius/40/39307_Inteligencia_artificial.pdf

Salas. (27 de julio de 2004). *Redes Neuronales Artificiales*. Obtenido de Universidad de Valparaíso Departamento de Computación:
https://www.academia.edu/24633757/Redes_Neuronales_Artificiales

Salazar, L. F. (2020). "Análisis de la calidad de agua del río Jerusalén efluente del río Pedro Carbo - Guayas para uso agrícola". Guayaquil, Ecuador. Obtenido de
<https://cia.uagraría.edu.ec/Archivos/SALAZAR%20HOLGUIN%20LUIS%20FAUTINI%20ANO.pdf>

Sánchez y Cabezas. (2018). Enfermedades infecciosas relacionadas con el agua en el Perú. *Rev. perú. med. exp. salud pública [online]*, 309-316. Obtenido de
<http://www.scielo.org.pe/pdf/rins/v35n2/a20v35n2.pdf>

Schlundt y Peter. (2014). *Una sola salud y seguridad alimentaria*. New York: National Food Institute, Technical University of Denmark (DTU), Copenhagen, Denmark. Obtenido de
https://link.springer.com/chapter/10.1007/978-4-431-55120-1_10

Science. (19 de Marzo de 2022). *Informe sobre el estado de la ciencia de datos de 2022*. Obtenido de
<https://www.coursera.org/articles/data-analytics>:
https://know.anaconda.com/rs/387-XNW-688/images/ANA_2022SODSReport.pdf

- Segura y Cols. (2021). Monitoreo de calidad de agua y predicción de coliformes fecales en playas de Montevideo mediante algoritmos de aprendizaje automático. *Innotec*, 1 - 15. Obtenido de <https://www.redalyc.org/journal/6061/606166107009/html/>
- Skerritt y Fletcher. (2015). *Professional Learning in Higher Education and Communities: Towards a New Vision for Action Research*. San diego: Macmillan Education palgrave. Obtenido de https://www.buscalibre.cl/libro-professional-learning-in-higher-education-and-communities-towards-a-new-vision-for-action-research/9781137455178/p/50875119?srsId=AfmBOoq2T8uJk75WXt7alxJE1iNHR5BYLR8MvIwK_M1_ro2rDkanZEqS
- Solano, A. M. (2020). "Monitoreo del índice de calidad de agua y su relación con la cafeína como contaminante emergente en la subcuenca del río Chumbao, Andahuaylas, Apurímac, Perú". Cusco, Perú. Obtenido de <https://repositorio.uandina.edu.pe/backend/api/core/bitstreams/5082c340-9132-4cfb-9983-82a69e928720/content>
- Solarte, Y., Peña , M., & Madera, C. (2006). Transmisión de protozoarios patógenos a través del agua para consumo humano. Colombia: Colombia Médica. Obtenido de https://www.redalyc.org/pdf/283/Resumenes/Resumen_28337111_1.pdf
- Sunass. (22 de Setiembre de 2022). *Determinación del Área de Prestación de Servicios del departamento de Apurímac*. Obtenido de Sunass: https://www.sunass.gob.pe/wp-content/uploads/2022/10/Informe-ADP-Apurimac_VP.pdf
- Tay, J., & Sánchez, J. (2002). Características de protozoarios y helmitos capaces de causar diarrea aguda en humanos. *Revista de Facultad de Medicina*. Obtenido de <https://www.medigraphic.com/pdfs/facmed/un-2002/un022e.pdf>

- Tecana, A. U. (2022). "Los niveles de investigación". Obtenido de <https://tauniversity.org/los-niveles-de-investigacion>
- Valdes y Moreno. (2011). Cólera: historia y actualidad. Revista de Ciencias. *Rev Ciencias Médicas vol.15 no.4 Pinar del Río oct.-dic. 2011*, 280 - 294. Obtenido de http://scielo.sld.cu/scielo.php?pid=S1561-31942011000400025&script=sci_arttext&tlng=en
- Vargas. (2020). "Calidad de agua por parámetros fisicoquímicos y microbiológicos, río Shilcayo sector Chontamuyo para proponer estrategias de conservación, San Martín - 2020". *Universidad Cesar Vallejo*, 31-35. Obtenido de <https://es.slideshare.net/slideshow/vargasmgsdpdf/263013735>
- Vargas, M. C. (2021). "Determinación de índice simplificado de calidad de agua en el río Chili, Arequipa 2019". Arequipa, Perú. Obtenido de <https://repositorio.unsa.edu.pe/server/api/core/bitstreams/5703e3a5-3a40-4b19-aadd-132968bb828c/content>
- Wang y Shanyou. (2023). *Predicción de PGA en alerta temprana de terremotos utilizando una red neuronal de memoria a corto plazo*. Kobe: Geophysical Journal International. Obtenido de <https://www.nature.com/articles/s41598-024-56004-6>
- Wendorff. (07 de Octubre de 2019). *Agua y fortalecimiento del núcleo familiar como base del desarrollo social y económico del Perú*. Obtenido de Ministro de Vivienda, Construcción y Saneamiento, : <https://library.fes.de/pdf-files/bueros/peru/20507.pdf>
- Yin. (2018). *Case Study Research and Applications: Design and Methods (6th ed.)*. Houston: Thousand Oaks CA. Obtenido de <https://uk.sagepub.com/en-gb/eur/case-study-research-and-applications/book250150>

Zunino M, E., & Pizarro P, R. (2007). Leptospirosis. *Puesta al día. Revista chilena de infectología*. Chile: <https://doi.org/10.4067/S0716-10182007000300008>.

/RV DQH[RV SDQHO IRWRJUiILFR \ RWURV GRFXPHQWRV F
GLJLWDO LQVWLWXFLRQDO HQ OD %LEOLRWHFD &HQWUDC