



**Ciencia Latina**  
Internacional

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.  
ISSN 2707-2207 / ISSN 2707-2215 (en línea), julio-agosto 2024,  
Volumen 8, Número 4.

[https://doi.org/10.37811/cl\\_rcm.v8i4](https://doi.org/10.37811/cl_rcm.v8i4)

**APLICACIÓN DE INTELIGENCIA ARTIFICIAL PARA  
LA OBTENCIÓN DE DATOS FALTANTES DE  
PRECIPITACIÓN PLUVIAL EN LA CUENCA RAVELO  
PARA LA CAPTACIÓN DE AGUA DULCE**

**APPLICATION OF ARTIFICIAL INTELLIGENCE FOR  
RETRIEVING MISSING RAINFALL DATA IN THE RAVELO  
BASIN FOR FRESHWATER COLLECTION**

**José Edgar Campos Serrano**

Universidad Mayor Real y Pontificia de San Francisco Xavier de Chuquisaca , Bolivia

**José Boris Bellido Santa María**

Universidad Mayor Real y Pontificia de San Francisco Xavier de Chuquisaca , Bolivia

**Carla Veronica Espada Aguilar**

Universidad Mayor Real y Pontificia de San Francisco Xavier de Chuquisaca , Bolivia

**Janeth Maciel Huaranca Callejas**

Universidad Mayor Real y Pontificia de San Francisco Xavier de Chuquisaca , Bolivia

**Alex Daniel Ibarra Tumiri**

Universidad Mayor Real y Pontificia de San Francisco Xavier de Chuquisaca , Bolivia

DOI: [https://doi.org/10.37811/cl\\_rcm.v8i4.12579](https://doi.org/10.37811/cl_rcm.v8i4.12579)

## Aplicación de Inteligencia Artificial para la Obtención de Datos Faltantes de Precipitación Pluvial en la Cuenca Ravelo para la Captación de Agua Dulce

**José Edgar Campos Serrano<sup>1</sup>**

[campos.edgar@usfx.bo](mailto:campos.edgar@usfx.bo)

<https://orcid.org/0009-0009-7682-7534>

Universidad Mayor Real y Pontificia de San Francisco Xavier de Chuquisaca Sucre, Bolivia

**José Boris Bellido Santa María**

[bellido.boris@usfx.bo](mailto:bellido.boris@usfx.bo)

<https://orcid.org/0000-0002-2380-7785>

Universidad Mayor, Real y Pontificia de San Francisco Xavier de Chuquisaca Sucre, Bolivia

**Carla Veronica Espada Aguilar**

[carlyveronica93@gmail.com](mailto:carlyveronica93@gmail.com)

<https://orcid.org/0009-0002-3078-7193>

Universidad Mayor, Real y Pontificia de San Francisco Xavier de Chuquisaca Sucre, Bolivia

**Janeth Maciel Huaranca Callejas**

[huaranca@gmail.com](mailto:huaranca@gmail.com)

<https://orcid.org/0009-0005-0905-0703>

Universidad Mayor Real y Pontificia de San Francisco Xavier de Chuquisaca Sucre, Bolivia

**Alex Daniel Ibarra Tumiri**

[sensey.dumb12@gmail.com](mailto:sensey.dumb12@gmail.com)

<https://orcid.org/0009-0002-0635-5972>

Universidad Mayor Real y Pontificia de San Francisco Xavier de Chuquisaca Sucre, Bolivia

### RESUMEN

El objetivo general es mejorar la obtención de datos faltantes de precipitación pluvial de la cuenta de Ravelo para la captación de agua dulce. Se sistematizaron métodos convencionales y de Inteligencia Artificial (IA) para imputar precipitaciones pluviales. Los métodos convencionales fueron: Servicio Meteorológico Nacional de EE.UU., Razón Normal y Regresión Lineal Simple. En IA, se aplicaron: Regresión Lineal Múltiple, Redes Neuronales Artificiales, Series Temporales, Árbol de Decisión, Perceptrón Multicapa, Bosques Aleatorios y Redes Neuronales Recurrentes, que requieren grandes volúmenes de datos. ELAPAS proporcionó datos de las estaciones de Tumpeka, Ravelo y Cajamarca (2018-2024). Se desarrollaron programas en Python para aplicar métodos convencionales e IA, usando los datos proporcionados para el entrenamiento y pruebas. La comparación de métodos se basó en el error cuadrático medio porcentual. La Regresión Lineal Simple resultó ser el mejor método convencional y el Árbol de Decisión fue el mejor método de IA. Los métodos convencionales dependen de estaciones meteorológicas cercanas, limitando su uso en áreas sin estas estaciones. Los métodos de IA, no requieren esta información y proporcionan estimaciones precisas con pocos años. Este estudio, realizado en la Cuenca de Ravelo, demostró que los modelos de IA son efectivos para estimar la precipitación pluvial.

**Palabras clave:** imputación, precipitación, hidrología, inteligencia artificial

---

<sup>1</sup> Autor principal.

Correspondencia: [bellido.boris@usfx.bo](mailto:bellido.boris@usfx.bo)

# Application of Artificial Intelligence for Retrieving Missing Rainfall Data in the Ravelo Basin for Freshwater Collection

## ABSTRACT

The general objective is to improve the retrieval of missing rainfall data from the Ravelo Basin for freshwater collection. Conventional and Artificial Intelligence (AI) methods were systematized to impute rainfall. The conventional methods were: U.S. National Weather Service, Normal Ratio, and Simple Linear Regression. In AI, Multiple Linear Regression, Artificial Neural Networks, Time Series, Decision Tree, Multilayer Perceptron, Random Forests, and Recurrent Neural Networks were applied, which require large volumes of data. ELAPAS provided data from the Tumpeka, Ravelo, and Cajamarca stations (2018-2024). Python programs were developed to apply conventional and AI methods, using the provided data for training and testing. The comparison of methods was based on the mean squared percentage error. Simple Linear Regression proved to be the best conventional method, and the Decision Tree was the best AI method. Conventional methods rely on nearby weather stations, limiting their use in areas without such stations. AI methods do not require this information and provide accurate estimates with just a few years of data. This study, conducted in the Ravelo Basin, demonstrated that AI models are effective in estimating rainfall.

**Keywords:** imputation, precipitation, hydrology, artificial intelligence

*Artículo recibido 14 junio 2024  
Aceptado para publicación: 17 julio 2024*



## INTRODUCCIÓN

La vida humana depende de la presencia de agua dulce. De los 1,4 BCM disponibles de los suministros de agua de nuestro planeta, el agua dulce representa menos del tres por ciento (Jumapam, 2018) como se muestra en la figura 1.

La mayor parte está ligada a los glaciares, los casquetes polares y los campos nevados, particularmente en la Antártida (Vereda, 2008). A diferencia de estos sectores particulares en el planeta, en nuestro entorno el ciclo del agua llega a tener un papel muy importante, entre sus tantas fases la que más destaca es la precipitación siendo la misma una de las fuentes de abastecimiento fundamentales, sobre todo en el la zona del altiplano y parte de los valles de Bolivia (Pacheco Apaza, 2016). Los datos de precipitación y temperatura son captados a través del uso de pluviómetros y estaciones meteorológicas terrestres. El Servicio Nacional de Meteorología e Hidrología (SENAMHI), entidad encargada de recopilar la información, cuenta con un total de 699 estaciones pluviométricas distribuidas en todo el país, de las cuales 547 se encuentran funcionando. (Ureña et al., 2018, p. 2). Su distribución se halla centrada en los ejes metropolitanos y sus alrededores, sin embargo, estos datos se recopilan manualmente y están sujetos a errores humanos, siendo una recolección aún más compleja la toma de datos en las zonas montañosas, (Huanca Aspi, 2019) alejadas y de difícil acceso. Debido a esto, el análisis de precipitaciones utilizando estaciones terrestres y de telemetría, presenta varios inconvenientes. (Araoz Rojas, 2018)

Por otro lado, la Empresa Local de Agua Potable y Alcantarillado Sucre (ELAPAS) cuenta con estaciones meteorológicas en diferentes cuencas que pueden servir para el suministro de agua potable para la ciudad de Sucre. (ELAPAS, Sucre, s. f.) En la cuenca Ravelo, se encuentran las estaciones de Ravelo, Cajamarca, Tumpeka y Potolo que proporcionan lecturas cada 15 minutos.

Como es el caso de la cuenca Ravelo, la cual se encuentra ubicada en una zona montañosa, lo que dificulta la obtención de datos y su accesibilidad. s (temporadas de lluvias). (Montero Torres, 2023) A causa de ese problema, la cuantificación del caudal genera incertidumbre en la toma de decisiones, para proyecciones futuras de estudios y proyectos hidráulicos e hidrológicos. (Caloir & Molina Carpio, 2015)



## Planteamiento del problema de Investigación

¿Cómo obtener datos faltantes de precipitación pluvial de la cuenta de Ravelo para la captación de agua dulce?

## Sustento Teórico

La presente investigación se basa en la integración de conocimientos de hidrología y de inteligencia artificial, con el objetivo de mejorar la captación de agua dulce en la cuenca Ravelo a través de la obtención de datos faltantes de precipitación pluvial. La combinación de ambas áreas de estudio permitiría desarrollar soluciones innovadoras y eficientes para la gestión hídrica en esta región.

## U.S.National Weather Service

Este procedimiento ha sido verificado teóricamente como empíricamente y considera que el dato faltante de una estación X por ejemplo, puede ser estimada en base a los datos observados en las estaciones circundantes. (Toro Trujillo et al., 2015)

El método puede ser aplicado para estimar valores diarios, mensuales o anuales faltantes. El método consiste en ponderar los valores observados en una cantidad W, igual al recíproco del cuadrado de la distancia entre cada estación vecina y la estación (Coras Merido, 2020)

Como se observa en la ecuación:

$$P_x = \frac{\sum P_i * w_i}{\sum w_i} \quad (1)$$

Donde:

$P_i$  = Precipitación observada para la fecha de la faltante, en las estaciones auxiliares circundantes.

$w_i = 1/D_i^2$  Siendo  $D_i$  la distancia entre cada estación circundante y la estación incompleta, en km.

## Razón normal

Este método se basa en la relación constante entre los valores de precipitación de dos estaciones, una de las cuales tiene datos completos y la otra no.

Los valores mensuales, anuales o medios tiende a ser constante. Para aplicar este método, se necesitan tres o más estaciones cercanas y confiables, que estén uniformemente espaciadas con respecto a la estación en estudio. (Pizarro et al., 2009)



$$P_x = \frac{N_x}{n} * \sum_{i=1}^n \frac{P_i}{N_i} \quad (2)$$

Donde:

$P_x$  = registro de precipitación a ser completado

$n$  = número de estaciones auxiliares

$P_i$  = precipitaciones de las estaciones vecinas

### **Regresión Lineal**

En el modelo de regresión lineal se requiere de una estación “donante” y preferiblemente cercana a la estación con datos faltantes.(Scheaffer, s. f.)

La fórmula es:

$$y = \beta_0 + \beta_1 * x \quad (3)$$

Donde:

$y$  = es el dato de precipitación a rellenar

$\beta_0$  y  $\beta_1$  = coeficientes de la regresión

$x$  = precipitaciones de la estación “donante”

Este método es mayormente usado cuando no se disponen de muchos datos de precipitación.

La inteligencia artificial se refiere al campo de estudio que busca desarrollar sistemas y algoritmos capaces de imitar el comportamiento humano, procesamiento de información y toma de decisiones de manera autónoma.(Moreno & Balcázar, 2023)

En el campo de la hidrología, la inteligencia artificial ha tenido una intervención significativa, utilizando técnicas como el aprendizaje automático, cómo las que se describen a continuación:

### **Regresión Lineal Múltiple**

La regresión lineal múltiple es un método estadístico utilizado para modelar la relación entre una variable dependiente y dos o más variables independientes. (Rojo & Abuín, s. f.)

En el contexto de la precipitación pluvial, este método se aplica para predecir la cantidad de lluvia utilizando múltiples variables ambientales como la temperatura, la humedad, la presión atmosférica.



### **Redes Neuronales Artificiales (RNA)**

Las redes neuronales artificiales son modelos computacionales inspirados en el funcionamiento del cerebro humano, utilizados para el aprendizaje automático y el procesamiento de datos. (Veintimilla & Cisneros, 2014)

En el caso de la precipitación pluvial, las RNAs pueden ser entrenadas para analizar múltiples variables ambientales y predecir la cantidad de lluvia con precisión.

### **Series Temporales**

Las series temporales son conjuntos de datos ordenados en función del tiempo. Estos datos se recopilan en intervalos regulares y se utilizan para analizar patrones, tendencias y comportamientos a lo largo del tiempo. (Corres et al., 2009)

En el contexto de la precipitación pluvial, las series temporales se emplean para estudiar y predecir la variabilidad de la lluvia en diferentes períodos.

### **Árbol de Decisión**

Un árbol de decisión es una estructura de árbol que representa un conjunto de decisiones y sus posibles consecuencias. En el aprendizaje automático, los árboles de decisión se utilizan para clasificar y predecir resultados en función de múltiples variables. (Hastie et al., 2009)

En el contexto de la precipitación pluvial, un árbol de decisión podría utilizarse para determinar las condiciones que influyen en la cantidad de lluvia en una región.

### **Perceptrón Multicapa**

El perceptrón multicapa es una arquitectura de red neuronal artificial que consta de múltiples capas de neuronas interconectadas, incluyendo una capa de entrada, una o más capas ocultas y una capa de salida. Este modelo es capaz de aprender y reconocer patrones complejos en los datos. (LeCun et al., 1989)

En el contexto de la precipitación pluvial, un perceptrón multicapa podría emplearse para modelar y predecir la cantidad de lluvia en función de diversas variables ambientales.

### **Bosques Aleatorios (Random Forests)**

Los bosques aleatorios son un conjunto de árboles de decisión que se utilizan para la clasificación y la regresión. Cada árbol en el bosque se entrena de forma independiente y la predicción final se obtiene



mediante la combinación de las predicciones de todos los árboles.(Medina Merino & Ñique Chacón, 2017)

En el contexto de la precipitación pluvial, los bosques aleatorios pueden utilizarse para predecir la cantidad de lluvia en función de múltiples variables ambientales.

### **Redes Neuronales Recurrentes(RNR)**

Las redes neuronales recurrentes (RNN) son modelos avanzados diseñados para procesar y analizar secuencias de datos temporales debido a su capacidad de mantener información de eventos pasados a través de sus conexiones recurrentes. (*Centeno Franco Alba TFG.pdf*, s. f.)

En la estimación de datos faltantes de precipitación pluvial, las RNN son especialmente útiles ya que pueden capturar las dependencias temporales en los datos históricos de precipitación.

**Objetivo General:** Mejorar la obtención de datos faltantes de precipitación pluvial de la cuenca de Ravelo para la captación de agua dulce.

### **Objetivos específicos:**

- Sistematizar los métodos hidrológicos convencionales para completar datos de precipitaciones.
- Sistematizar los métodos de Inteligencia Artificial para completar datos.
- Caracterizar la cuenca de estudio con los datos proporcionados de la Empresa ELAPAS y visitas In-Situ.
- Aplicar modelos hidrológicos para completar datos faltantes de precipitaciones de los años 2018-2023.
- Aplicar modelos de Inteligencia Artificial para completar datos faltantes de precipitaciones de los años 2018-2023.
- Comparar los resultados obtenidos para verificar su validez.

## **METODOLOGÍA**

### **1. Caracterizar la cuenca de estudio con los datos proporcionados por ELAPAS y visitas In-Situ**

Para caracterizar la cuenca de estudio se utilizó el Dron para obtener una información mucho más detallada de la zona como se muestra en la Figura 2.



### **Cobertura Vegetal y uso de suelo**

En los diferentes pisos altitudinales (alto, medio y bajo), se distingue diferente cobertura vegetal como pudimos observar en las visitas realizadas; así por ejemplo, en la parte alta (3800 a 4000 msnm) predomina la formación de pastizal con poca presencia de cultivos y la de pajonal como se observa en la Figura 4. La parte media (3500 a 3800 msnm) es ocupada por matorral de porte bajo, cultivos anuales a secano en las planicies y pequeñas manchas de especies forestales. (Revista de SIGEP)

Finalmente en las partes bajas partes bajas (3500 a 2300 msnm), predominan los cultivos cercanos en las laderas y cultivos bajo riego en los márgenes del río como se muestra en la figura 3. También se encuentran sectores con especies introducidas de árboles, como pinos y eucaliptos. En sectores se observa la presencia de molles, sauces y otras especies nativas.

### **Análisis hidrológico**

Los análisis y procesamiento de los registros de precipitación en estaciones dentro y cerca del área de las cuencas, ha permitido establecer las precipitaciones promedio en la cuenca, de donde se observa que la cuenca Ravelo hasta la Toma Existente, es la que mayor precipitación recibe a nivel anual y mensual, en tanto que la cuenca conjunta de los ríos Potolo y Ravelo hasta el sitio de la Toma Nueva Proyectada, tiene una precipitación promedio anual de unos 930,1 mm.

En toda el área, la época de lluvias abarca los meses de noviembre a marzo, con octubre y abril como periodo de transición, en tanto que los meses de mayo a septiembre constituyen el periodo de estiaje, con apenas un 5% del total anual.

### **Obtener los datos de ELAPAS**

Para la investigación se solicitó la información necesaria de la empresa ELAPAS gracias a solicitudes y los convenios existentes entre la universidad y la empresa, se obtuvo información de los datos de las estaciones de Ravelo, Tumpeka y Cajamarca, como se observa en la Tabla 1.

*Obtención de los parámetros de la cuenca:* Para realizar un estudio hidrológico de la cuenca, fue necesario la recopilación de información acerca de la geografía de la cuenca, como su área, tipo de suelo, vegetación. Datos que fueron obtenidos gracias a la empresa ELAPAS.

*Recolección de los datos meteorológicos:* Es fundamental recopilar datos sobre la precipitación, temperatura, etc. y cualquier otro dato meteorológico relevante para la cuenca en estudio. Estos datos también fueron brindados por la empresa ELAPAS, con las siguientes variables:

- **TIMESTAMP:** Fecha y hora del registro de la lectura.
- **RECORD:** Número incremental con el número de la lectura.
- **PBar:** Medida de la presión barométrica.
- **PrecipP:** Medición de la precipitación pluvial.
- **DirV:** Dirección de viento.
- **RH:** Humedad relativa.
- **TA:** Temperatura ambiente.
- **VelV:** Velocidad del viento.
- **ET:** Evapotranspiración.
- **R-Rad:** Radiación.
- **Rso:** No se conoce el significado de la variable.
- **Velv\_TMn:** No se conoce el significado de la variable.

### **Visita de campo**

Con el uso del Dron se pudo obtener mayor detalle en la caracterización de la cuenca, Al emplear técnicas de fotogrametría y teledetección se puede obtener información detallada sobre el uso de suelo y cobertura vegetal de la cuenca, lo que facilitó la caracterización de la misma.

## **2. Limpieza de los datos**

Antes de poder determinar los métodos es necesario un preprocesamiento y análisis de los datos, para ello se creó un código de python utilizando la librería Pandas.

El código lee los datos, elimina los datos que no sean relevantes en la precipitación y elimina las lecturas que tengan datos faltantes, también se eligió un rango de fechas para el entrenamiento y otro para las pruebas.

## **3. Determinar las variables que influyen en la precipitación**

Una vez los datos hayan sido procesados se procederá a realizar la correlación entre la precipitación y las distintas variables, para ello se usa la correlación de Pearson, como se puede ver en la tabla 2.



#### **4. Obtención de las distancias entre estaciones**

Mediante el uso de Google Earth se obtuvo las distancias entre las diferentes estaciones y la estación de Tumpeka, estas distancias son necesarias para algunos métodos hidrológicos convencionales

#### **5. Aplicar los métodos hidrológicos convencionales**

Para aplicar los métodos convencionales, se recopiló información completa de los datos reales de las estaciones meteorológicas de Ravelo, Tumpeka y Cajamarca, identificando los meses más lluviosos. Posteriormente, se eliminaron los datos reales de los meses de noviembre y diciembre de 2022, así como de enero y febrero de 2023, para luego ser estimados utilizando los siguientes métodos de estimación: Método del Servicio Meteorológico Nacional de EE.UU., Método de la Razón Normal y el Método de Regresión Lineal.

#### **6. Aplicar los métodos de Inteligencia Artificial**

Para utilizar los métodos de Inteligencia Artificial, primero se obtuvo información pluviométrica y meteorológica de la estación de Tumpeka, con mediciones cada 15 minutos desde el año 2018 hasta el 2024. Posteriormente, se utilizó la biblioteca Pandas para manejar los más de 200,000 datos recopilados. Se realizó un filtrado de datos, limpiando las columnas sin información, y se llevó a cabo una correlación para identificar las variables más adecuadas y óptimas para el entrenamiento. Las variables seleccionadas fueron: presión barométrica, humedad relativa, velocidad del viento y dirección del viento. Luego, se identificaron los rangos de fechas para el entrenamiento y la prueba, siendo estos:

Fechas de entrenamiento: 2018-01-01 al 2022-10-31, y 2023-03-01, 2024-01-31

Fechas de prueba: 2022-11-01 al 2023-02-28

Estos rangos se aplicaron a los métodos de inteligencia artificial, tales como regresión lineal múltiple, redes neuronales artificiales, series temporales, árbol de decisión, perceptrón multicapa, bosques aleatorios y redes neuronales recurrentes, mismos que se encuentra implementados en la biblioteca sklearn.

## 7. Comparar los resultados de los métodos hidrológicos y de los métodos de Inteligencia Artificial.

Para realizar la comparación se aplicó el error porcentual:

1. Calcular el Error Cuadrático Medio (ECM):

$$ECM = \frac{1}{n} \sum_{i=1}^n (y_{real_i} - y_{pred_i})^2$$

donde:

$y_{real_i}$  son los valores reales

$y_{pred_i}$  son los valores predichos.

2. Calcular la Raíz del Error Cuadrático Medio (RECM):

$$RECM = \sqrt{ECM}$$

3. Calcular el rango de los valores reales:

$$Rango = \max(y_{real_i}) - \min(y_{real_i})$$

4. Convertir el RECM en porcentaje del rango:

$$ECM\% = \frac{RECM}{Rango} \times 100\%$$

## RESULTADOS Y DISCUSIÓN

Después de aplicar los métodos hidrológicos convencionales y de los métodos de Inteligencia Artificial se obtuvieron los resultados de la tabla 3, donde se puede observar el error cuadrático medio porcentual (ECM%) de cada método.

El uso del Método del Servicio Meteorológico Nacional de EE.UU. proporciona estimaciones razonables para los datos faltantes. Las diferencias entre los valores originales y estimados son esperadas debido a la combinación ponderada de los datos de la estaciones de Ravelo y Cajamarca, reflejando una media basada en la proximidad de las estaciones vecinas. La variabilidad climática y las diferencias en las condiciones locales también contribuyen a estas diferencias. La inclusión de una o más estaciones adicionales en un radio de 20 km, proporcionará una muestra más amplia y diversa,

permitiendo una representación más precisa de las condiciones climáticas de la región y reduciendo el sesgo presente al usar solo dos estaciones.

La razón normal es uno de los métodos más utilizados en el relleno de datos según los datos de error obtenidos es el segundo más eficaz de los métodos convencionales con un valor de 6,62% . Algunas razones por la que el método puede variar son: en primer lugar, el número de estaciones de referencia: se deben seleccionar estaciones de referencia que sean representativas de la región en la que se está trabajando y que tengan datos suficientes para calcular las razones normales. En segundo lugar, la variabilidad climática: factores como el cambio climático pueden afectar la validez de las razones normales, por lo que es importante tener en cuenta la variabilidad climática al aplicar este método.

La Regresión Lineal Simple estima un error del 1,54% indica una alta precisión en la predicción de precipitaciones. Este bajo nivel de error sugiere que el modelo de regresión lineal utilizado es capaz de capturar de manera efectiva las relaciones lineales entre las variables involucradas en la predicción, esto es importante en el contexto de la meteorología, donde incluso pequeñas desviaciones en las predicciones pueden tener un impacto significativo en la toma de decisiones y la planificación de actividades sensibles al clima. Si las condiciones meteorológicas entre estas estaciones son significativamente diferentes, esto puede llevar a una mayor discrepancia entre las predicciones y las observaciones reales. Por lo tanto, es fundamental considerar la representatividad de las estaciones cercanas al realizar predicciones meteorológicas basadas en modelos estadísticos como la regresión lineal.

El uso de la regresión lineal múltiple para la predicción de precipitaciones ha mostrado un error porcentual del 2,23%. Este método, que modela la relación entre una variable dependiente y múltiples variables independientes, es eficaz para capturar relaciones lineales entre los datos, proporcionando predicciones razonablemente precisas. Sin embargo, la regresión lineal múltiple puede ser limitada en su capacidad para modelar relaciones no lineales complejas presentes en los datos. Para asegurar un rendimiento consistente y robusto, es crucial prestar atención a la selección de variables y a la calidad de los datos, así como utilizar técnicas de validación cruzada para evaluar y mejorar la generalización del modelo.



El uso de redes neuronales artificiales para la predicción de precipitaciones ha mostrado resultados prometedores, con un error porcentual del 2,93%. Este bajo nivel de error destaca la eficacia de las redes neuronales artificiales, en capturar relaciones complejas en los datos y su capacidad para proporcionar predicciones precisas, sin embargo este valor es significativo en comparación a los resultados de los demás métodos por lo que se recomienda ajustar los hiperparámetros del modelo, normalizar los datos de entrada, aumentar la cantidad de datos de entrenamiento, para lograr una mejor generalización y reducir el riesgo de sobreajuste.

El uso de técnicas de predicción basadas en series temporales para la estimación de precipitaciones ha mostrado un error porcentual del 2,61%, destacando su eficacia en capturar patrones dependientes del tiempo. Estos modelos son adecuados para datos secuenciales y proporcionan predicciones precisas. Para mantener un rendimiento consistente y robusto, es crucial asegurar la calidad de los datos, ajustar adecuadamente el modelo y utilizar validación cruzada temporal. Incorporar factores exógenos relevantes y un adecuado preprocesamiento de datos también puede mejorar la precisión.

El método de regresión basado en un árbol de decisión, configurado con una profundidad máxima de 5, mostró un desempeño razonable al predecir la precipitación total, con un error cuadrático medio relativo al rango (RECM%) de 1.84%. Esto indica una desviación mínima respecto al rango total de los valores reales de precipitación. Aunque el modelo es interpretable y captura relaciones no lineales, existe el riesgo de sobreajuste y sensibilidad a variaciones en los datos. Aunque el modelo es interpretable y captura relaciones no lineales, existe el riesgo de sobreajuste y sensibilidad a variaciones en los datos.

El uso del Perceptrón Multicapa para la predicción de precipitaciones ha mostrado un error porcentual del 3,87%. Aunque este método tiene la capacidad de capturar relaciones no lineales en los datos, el mayor nivel de error en comparación con otros modelos sugiere que puede no ser el más adecuado para este conjunto de datos específico, se estima que el modelo MLP no está capturando adecuadamente la relación entre las variables independientes y la variable objetivo. Para mejorar el modelo, se pueden ajustar los hiperparámetros, normalizar los datos de entrada, aumentar la cantidad de datos de entrenamiento y explorar otras arquitecturas de redes neuronales según el tipo de datos y el problema a resolver.



El uso de Bosques Aleatorios para la predicción de precipitaciones ha mostrado resultados sobresalientes, con un error porcentual del 1,94%. Este bajo nivel de error destaca la capacidad de los Bosques Aleatorios para manejar complejidades en los datos sin sobreajustar, proporcionando predicciones precisas y confiables. Este valor relativamente bajo sugiere que el modelo tiene una buena capacidad para hacer predicciones precisas en este conjunto de datos específico. Las ventajas del método de Bosques Aleatorios incluyen su capacidad para manejar conjuntos de datos grandes con muchas características, su resistencia al sobreajuste gracias a la combinación de múltiples árboles, y su capacidad para manejar datos faltantes y variables categóricas sin necesidad de preprocesamiento adicional.

El uso de Redes Neuronales Recurrentes para la predicción de precipitaciones ha mostrado un error porcentual del 2,19%. Este resultado indica que las RNNs son eficaces en capturar dependencias temporales en los datos, proporcionando predicciones precisas. Para asegurar un rendimiento consistente y robusto, es crucial optimizar la arquitectura de la red, ajustar los hiperparámetros y mejorar la calidad de los datos. Utilizar técnicas de validación cruzada también es esencial para evaluar y mejorar la generalización del modelo.

## ILUSTRACIONES, TABLAS, FIGURAS

**Tabla 1.:** Valores de precipitaciones de las estaciones meteorológicas.

Estación	Registros meteorológicos (Cada 15 minutos)
Cajamarca	109 329
Tumpeka	110 134
Ravelo	145 131

Fuente: Elaboración Propia

**Tabla 2.:** Correlación de variables para la precipitación.

VARIABLES	Correlación
Humedad Relativa	0,110408
Velocidad del viento	0,011158
Dirección del viento	0,009436
Presión Barométrica	0,005318

Fuente: Elaboración Propia



**Tabla 3.:** Distancia entre la estación de Tumpeka y otras estaciones.

Estación	Distancia (km)
Potolo	10,730
Cajamarca	5,890
Ravelo	18,445

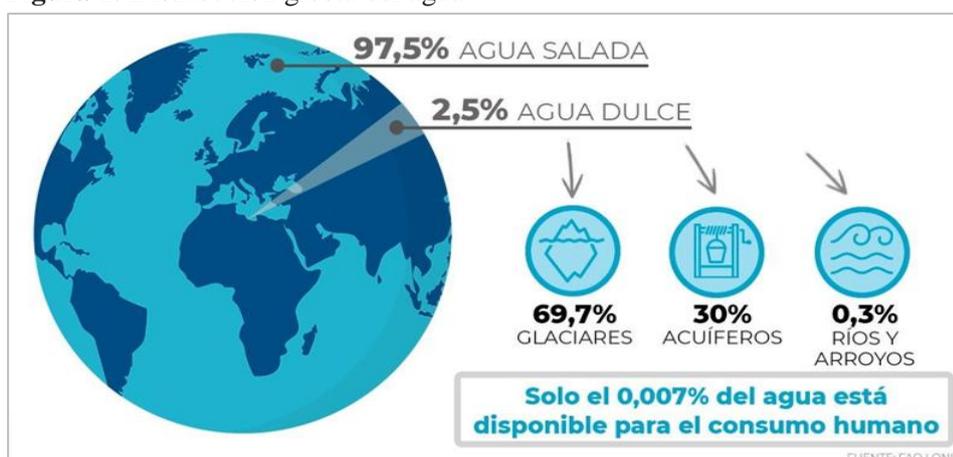
Fuente: Elaboración propia

**Tabla 4.:** Error cuadrático medio porcentual (ECM%) de cada método.

Métodos utilizados	Tipo	ECM %
Servicio Meteorológico Nacional de EE.UU.	Hidrológico convencional	31,78
Razón Normal	Hidrológico convencional	6,62
Regresión Lineal Simple	Hidrológico convencional	1,54
Regresión Lineal Múltiple	Inteligencia Artificial	2,23
Redes Neuronales Artificiales	Inteligencia Artificial	2,93
Series Temporales	Inteligencia Artificial	2,61
Árbol de Decisión	Inteligencia Artificial	1,84 %
Perceptrón Multicapa	Inteligencia Artificial	3,87 %
Bosques Aleatorios	Inteligencia Artificial	1,94 %
Redes neuronales Recurrentes	Inteligencia Artificial	2,19 %

Fuente: Elaboración Propia

**Figura 1:** Distribución global del agua



Fuente: (Distribución de Agua en el Planeta | Jumapam, s. f., p. 20)

**Figura 2.:** Vista panoramica de la obra de toma vista desde el Dron



Fuente: Elaboración Propia

**Figura 3.:** Cobertura Vegetal de la zona



Fuente: Elaboración Propia

**Figura 4.:** Cobertura Vegetal de la zona



Fuente: Elaboración Propia

## CONCLUSIONES

Los métodos convencionales utilizados fueron: Servicio Meteorológico Nacional de EE.UU., Razón Normal y Regresión Lineal Simple, los mismos que permiten completar los datos faltantes de precipitaciones pluviales a partir de la información de otras estaciones.

Los métodos de Inteligencia Artificial que pueden ser usados en la predicción de datos de precipitación pluvial fueron: Regresión Lineal Múltiple, Redes Neuronales Artificiales, Series Temporales, Árbol de Decisión, Perceptrón Multicapa, Bosques Aleatorios y Redes neuronales Recurrentes, los cuales requieren grandes volúmenes de datos para encontrar un modelo que permita realizar la predicción de manera adecuada.

Se realizaron las visitas a campo, en donde se observaron los pluviómetros cercanos a la zona de estudio, los pluviómetros que se encuentran en funcionamiento, se observaron los manejos y manipulación por el personal, también se observó el tipo de cobertura vegetal con la que cuenta el lugar. Por otro lado, la empresa ELAPAS proporcionó datos históricos de las estaciones de Tumpeka, Ravelo y Cajamarca de precipitación desde 2018 a 2024 con las variables: Fecha y hora del registro de la lectura, Número incremental con el número de la lectura, Medida de la presión barométrica, Medición de la precipitación pluvial, Dirección de viento, Humedad relativa, Temperatura ambiente, Velocidad del viento, Evapotranspiración, Radiación.

Se desarrollaron programas usando python que implementan los métodos hidrológicos convencionales, dicho programas permiten la aplicación rápida y efectiva de los métodos sobre un conjunto de datos de precipitaciones pluviales. De los meses de noviembre de 2022, diciembre de 2022, enero de 2023 y febrero de 2023.

Con sklearn se desarrollaron programas usando python que permiten realizar el entrenamiento, predicción y cálculo de errores de todos los métodos de Inteligencia Artificial estudiados. Se utilizaron los datos de los meses de noviembre de 2022, diciembre de 2022, enero de 2023 y febrero de 2023, como datos de pruebas; con los otros datos se procedió a entrenar cada método.

Para poder realizar la comparación entre los diferentes métodos se utilizó el error cuadrático medio porcentual, dando como resultado a la Regresión Lineal Simple como el mejor método hidrológico convencional con un error del 1,54 %. Por otro lado, el método de Árbol de Decisión es el mejor método



de Inteligencia Artificial con un error de 1,84 % y en segundo lugar el método de Bosques Aleatorios con un error de 1,94 %. La diferencia de error entre los métodos es tan pequeña que no se puede determinar claramente cual es el mejor de todos.

Finalmente, los métodos hidrológicos tradicionales dependen de estaciones meteorológicas cercanas para completar datos faltantes, lo que limita su utilidad en áreas sin estas estaciones. En contraste, los métodos basados en Inteligencia Artificial, como los modelos de Bosques Aleatorios y Árbol de Decisión, no requieren esta información externa y logran estimaciones precisas de datos faltantes de precipitación pluvial, incluso con pocos años de información disponible, pero con una gran cantidad de datos de las precipitaciones en dichos años. Estos métodos consideran variables dependientes como dirección del viento, presión barométrica, velocidad del viento y humedad relativa, lo que resalta su capacidad para proporcionar resultados efectivos y confiables en condiciones de datos limitados. El estudio se llevó a cabo en la estación de Tumpeka, la obra de toma de agua potable de Sucre, utilizando datos de una estación meteorológica ubicada en el mismo lugar. A pesar de la limitación en la disponibilidad de años de información, los modelos de Inteligencia Artificial demostraron su eficacia al proporcionar estimaciones precisas de precipitación pluvial, mostrando su utilidad incluso en ubicaciones estratégicas como la estación de Tumpeka.

## REFERENCIAS BIBLIOGRAFICAS

1. Araoz Rojas, J. (2018). Elaboración de un Plan de Seguridad del Agua PSA, para el aprovechamiento de las fuentes hídricas que abastecen los municipios de Dibulla y Riohacha. *Ingeniería Civil*. [https://ciencia.lasalle.edu.co/ing\\_civil/530](https://ciencia.lasalle.edu.co/ing_civil/530)
2. Caloir, B., & Molina Carpio, J. (2015). *Modelo de gestión de la cuenca Mauri – Desaguadero y escenarios de cambio climático. Informe final*. <https://doi.org/10.13140/RG.2.2.18203.89122>
3. Coras Merido, P. (2020). *Drenaje Agrícola*.
4. Corres, G., Esteban, A., García, J. C., & Zárate, C. (2009). Análisis de series temporales. *Revista Ingeniería Industrial*, 8(1), Article 1. <https://revistas.ubiobio.cl/index.php/RI/article/view/80>
5. *ELAPAS - Empresa Local de Agua Potable y Alcantarrillado Sucre*. (s. f.). Recuperado 25 de julio de 2024, de <https://web.elapas.com.bo/contenido.php?id=8>



6. Huanca Aspi, N. (2019). *Uso de tecnologías de información geográfica para el modelamiento de riesgos de inundación en la canalización del río Huayllani—Zona de Achumani, La Paz* [Thesis]. <http://repositorio.umsa.bo/xmlui/handle/123456789/33349>
7. *Inicio—Jumapam.gob.mx*. (s. f.). Recuperado 25 de julio de 2024, de <https://jumapam.gob.mx/>
8. Medina Merino, R. F., & Ñique Chacón, C. I. (2017). Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *Interfases*, 10, 165-189.
9. Montero Torres, J. (2023). Importancia de las cuencas Cachi Mayu y Cajamarca en Sucre: Una revisión basada en ecorregiones, políticas y servicios ambientales. *Revista de Investigación e Innovación Agropecuaria y de Recursos Naturales*, 10(3), 106-125.  
<https://doi.org/10.53287/mgso3359ju99y>
10. Moreno, E. G., & Balcázar, M. del C. S. (2023). EFECTOS DE LA APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA CONTABILIDAD Y LA TOMA DE DECISIONES. *GESTIÓN*, 1(1), Article 1. <https://revistap.ejeutap.edu.co/index.php/Gestion/article/view/71>
11. Pacheco Apaza, R. R. (2016). *Análisis de la disponibilidad de agua en el crecimiento económico de Bolivia: Periodo 1990 - 2014* [Thesis].  
<http://repositorio.umsa.bo/xmlui/handle/123456789/9952>
12. Pizarro, R., Ausensi, P., Aravena, D., Sangüesa, C., León, L., & Balocchi, F. (2009). Evaluación de Métodos Hidrológicos para la Completación de datos faltantes de precipitación en estaciones de la Región del Maule, Chile. *Aqua-LAC*, 1, 172-184.  
<https://doi.org/10.29104/phi-aqualac/2009-v1-2-07>
13. Rojo, J. M., & Abuín, J. M. R. (s. f.). *Regresión lineal múltiple*.
14. Scheaffer, R. L. (s. f.). *Elementary Survey Sampling*.
15. Toro Trujillo, A. M., Arteaga Ramírez, R., Vázquez Peña, M. A., & Ibáñez Castillo, L. A. (2015). Relleno de series diarias de precipitación, temperatura mínima, máxima de la región norte del Urabá Antioqueño. *Revista mexicana de ciencias agrícolas*, 6(3), 577-588.
16. Ureña, J. E., Vallejos, A. G., Saavedra, O. C., & Escalera, A. C. (2018). EVALUACIÓN DE LA PRECIPITACIÓN DISTRIBUIDA EN LA CUENCA KATARI BASADO EN



TECNOLOGÍA SATELITAL Y PRODUCTOS DERIVADOS. *Investigación & Desarrollo*, 18(1), 35-51.

17. Veintimilla, J. E., & Cisneros, F. (2014). Redes Neuronales Artificiales (RNA) aplicadas en la prediccion de caudales para intervalos de tiempo horarios. *Revista Tecnológica - ESPOL*, 27(2), Article 2. <https://rte.espol.edu.ec/index.php/tecnologica/article/view/304>
18. Vereda, M. (2008). Tierra del Fuego y Antártica: Un inventario de recursos turísticos desde la idea de complementariedad de destino. *Estudios y perspectivas en turismo*, 17(3), 199-225.
19. Hastie, T., Tibshirani, R., & Friedman, J. (2009b). Neural Networks. En T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (pp. 389-416). Springer. [https://doi.org/10.1007/978-0-387-84858-7\\_11](https://doi.org/10.1007/978-0-387-84858-7_11)
20. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten Digit Recognition with a Back-Propagation Network. *Advances in Neural Information Processing Systems*, 2. <https://proceedings.neurips.cc/paper/1989/hash/53c3bce66e43be4f209556518c2fcb54-Abstract.html>

