



# **INTRODUCCIÓN A MACHINE LEARNING PARA PRONÓSTICO HIDROLÓGICO**

Paul Muñoz

Abril 19, 2025

# ¿QUÉ ES MACHINE LEARNING?

- *Aprendizaje Automático*
- ML consiste en programar computadoras para que aprendan a partir de datos
- ML permite a las computadoras aprender sin ser programadas explícitamente
- Los nuevos desarrollos permiten que ML simule condiciones no vistas o extremas

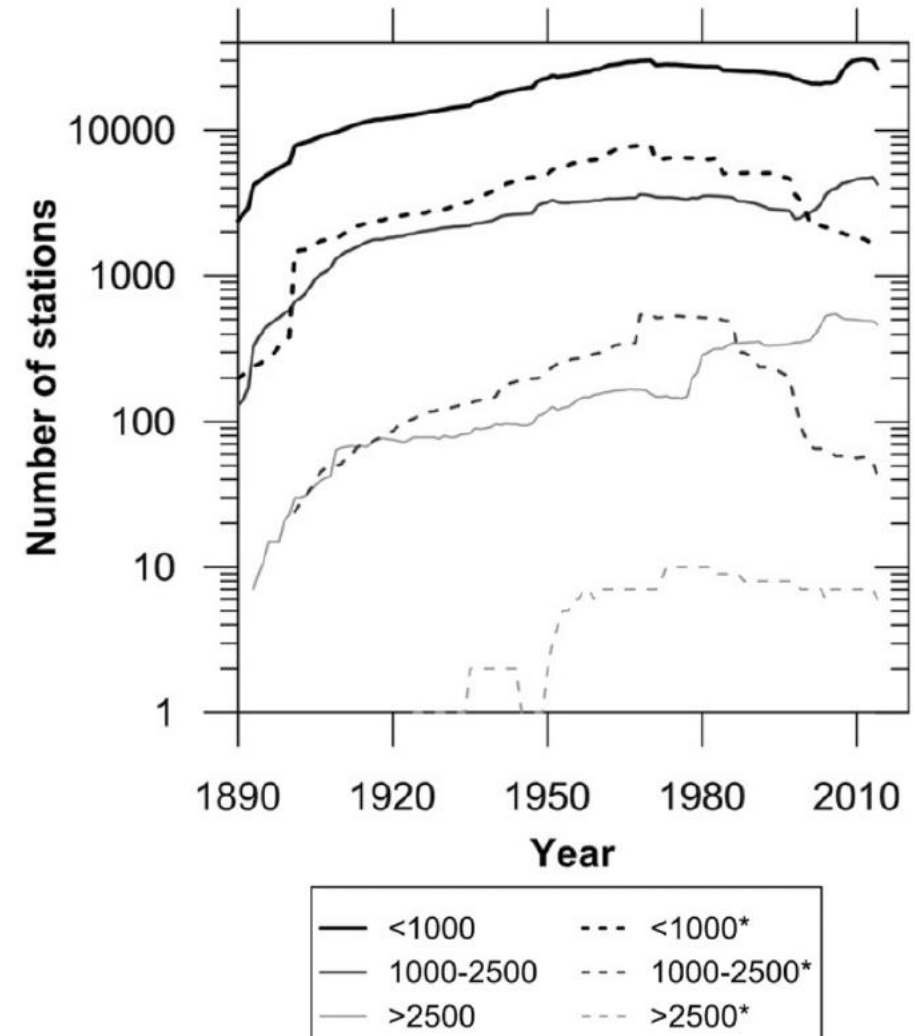
# ¿POR QUÉ MACHINE LEARNING?

- En las últimas décadas, el uso de técnicas de Machine Learning (ML) en la ingeniería, hidrología y meteorología ha crecido exponencialmente.
- El uso de ML en el pronóstico de caudales puede mejorar la precisión al identificar patrones y relaciones no obvias en los datos.
- Los avances tecnológicos (sensores, mediciones) en hidrología y climatología han permitido una mayor precisión en las predicciones hidrológicas.

# TRES BUENAS RAZONES PARA USAR ML

## 1. Cantidad de datos actuales y su crecimiento

- Monitoreo hidrometeorológico in situ  
Sensores de bajo costo e IoT
- Observación de la Tierra  
USGS y ESA producen ~1,5 TB por día
- Proyecciones de cambio climático  
Múltiples escenarios y experimentos

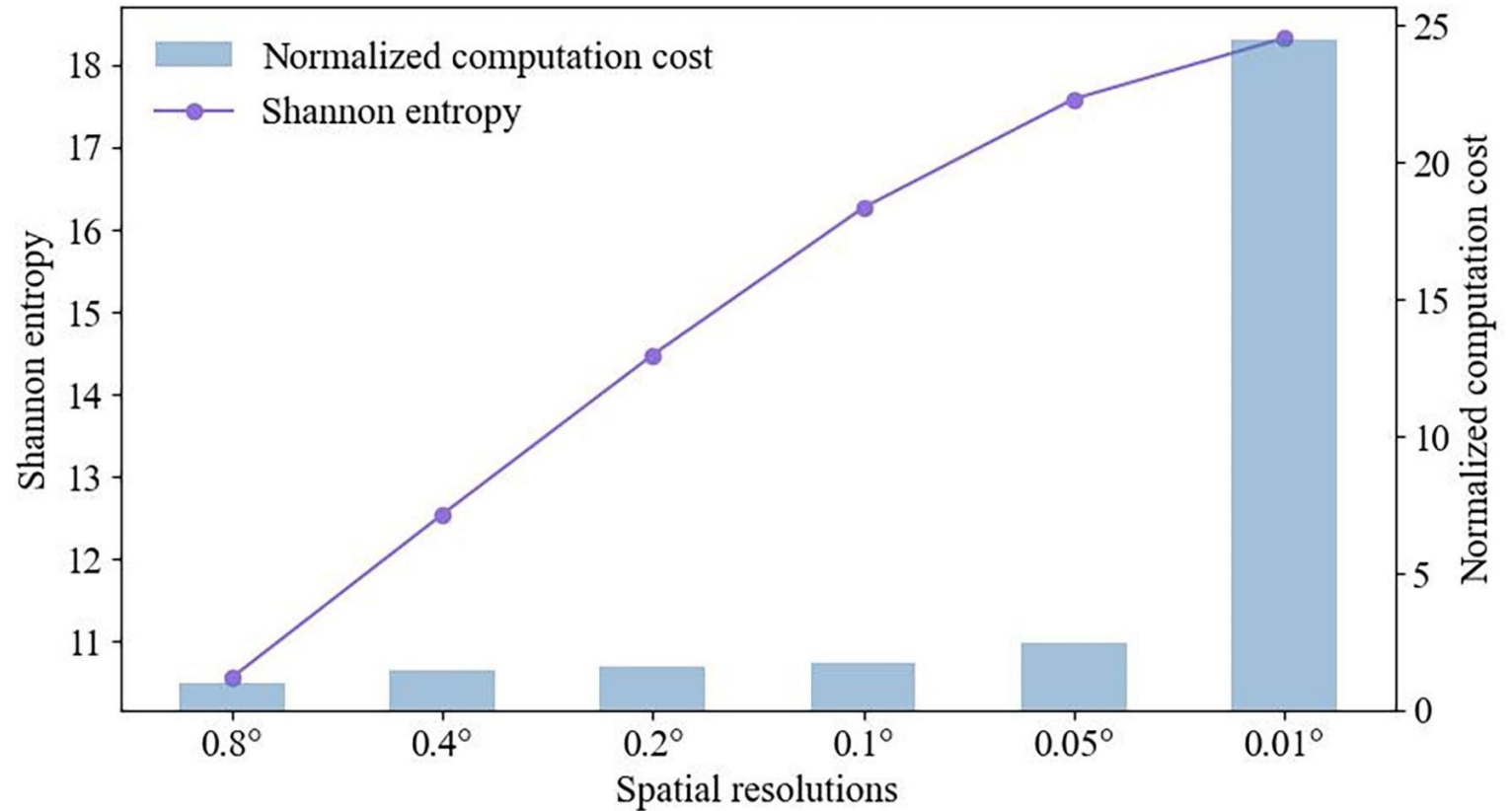


# TRES BUENAS RAZONES PARA USAR ML

## 2. Creciente? **poder computacional**

Trade-off:

Información y costo computacional



# TRES BUENAS RAZONES PARA USAR ML

## 3. Los algoritmos de ML se vuelven cada vez más complejos/inteligentes

- Problemas espaciales
- Problemas de series de tiempo
- Transferencia de modelos
- Optimización
- ...

# IMPACTO DEL ML EN LA CIENCIA

Desde 2020

Journal of Hydrology | Supports open access

Submit your article ↗

Articles & Issues ▾ About ▾ Publish ▾ Order journal ↗

🔍 Search in this journal

Guide for authors

Articles

Latest published

Articles in press

Top cited

Most downloaded

Most popular

The most cited articles published since January 2022. Source: [Scopus](#) ↗

Research article ○ Abstract only

Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation

Yuanhao Xu, ... Shuli Wang


May 2022

Research article ● Open access

Improving streamflow prediction in the WRF-Hydro model with LSTM networks

Kyeungwoo Cho, Yeonjoo Kim

February 2022

 [View PDF](#)

Research article ○ Abstract only

Do ERA5 and ERA5-land precipitation estimates outperform satellite-based precipitation products? A comprehensive comparison between state-of-the-art model-based and satellite-based precipitation products over mainland China

Jintao Xu, ... Jie Peng

February 2022

Research article ○ Abstract only

Runoff changes in the major river basins of China and their responses to potential driving forces

Lin Yang, ... Yunpeng Bai

April 2022

Research article ○ Abstract only

Study on optimization and combination strategy of multiple daily runoff prediction models coupled with physical mechanism and LSTM

Jun Guo, ... Hairong Zhang

September 2023

Research article ○ Abstract only

Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN

Fatemeh Rezaei Aderyani, ... Fatemeh Jafari

November 2022

Research article ○ Abstract only

Prediction of estuarine water quality using interpretable machine learning approach

Shuo Wang, ... Shengkang Liang

February 2022

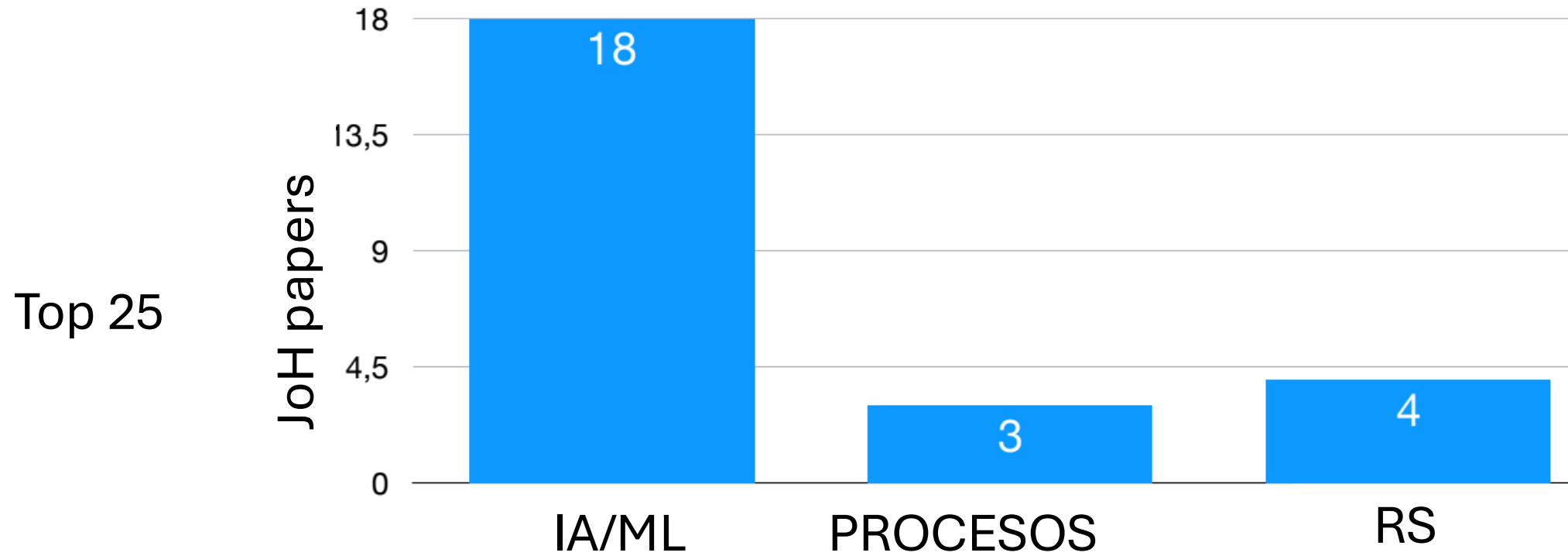
Research article ○ Abstract only

Stacked machine learning algorithms and bidirectional long short-term memory networks for multi-step ahead streamflow forecasting: A comparative study

Francesco Granata, ... Giovanni de Marinis

October 2022

# IMPACTO DEL ML EN LA CIENCIA





# TAREAS DE ML

Aprendizaje supervisado

▼ REGRESIÓN

Valores numéricos

Meteorology and Atmospheric Physics (2025) 137:5  
<https://doi.org/10.1007/s00703-024-01053-9>

ORIGINAL PAPER

Precipitation forecasting using random forest over an ecuadorian andes basin

Martín Montenegro<sup>1</sup> · Rolando Céleri<sup>1,2</sup> · Johanna Orellana-Alvear<sup>1</sup> · Paúl Muñoz<sup>1,2</sup> · Mario Córdova<sup>1,2</sup>

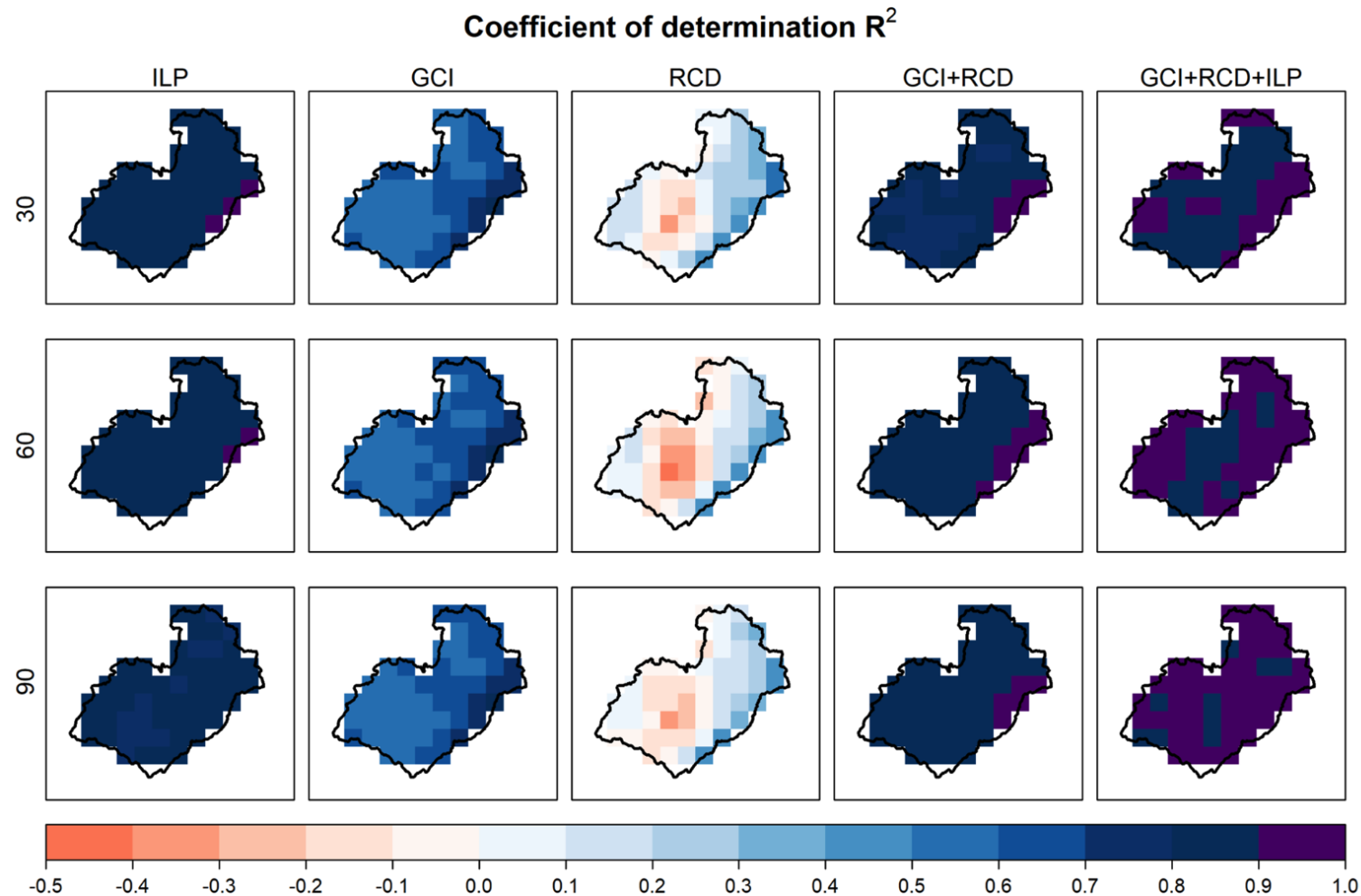


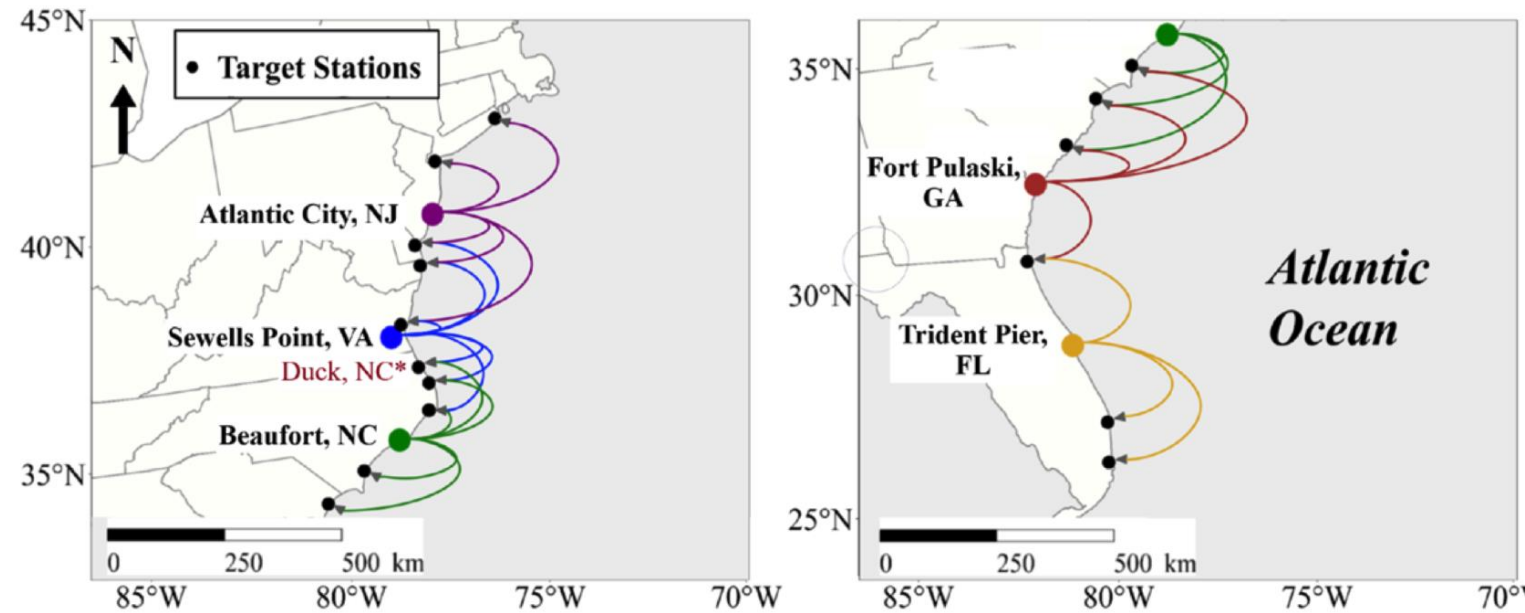
Fig. 4 Determination coefficient  $R^2$  in each box over the Paute basin. In columns the models and in rows the forecasting time steps

# TAREAS DE ML

## Aprendizaje supervisado

▼ REGRESIÓN  
Valores numéricos

(Transfer models from training to target stations)



### Water Resources Research<sup>\*</sup>

RESEARCH ARTICLE  
10.1029/2024WR039054

Special Collection:  
Forcing, response, and impacts of  
coastal storms in a changing  
climate

Predicting the Evolution of Extreme Water Levels With  
Long Short-Term Memory Station-Based Approximated  
Models and Transfer Learning Techniques

Samuel Daramola<sup>1</sup> , David F. Muñoz<sup>1</sup> , Paul Muñoz<sup>2</sup> , Siddharth Saxena<sup>1</sup> , and  
Jennifer Irish<sup>1</sup> 



# TAREAS DE ML

## Aprendizaje supervisado

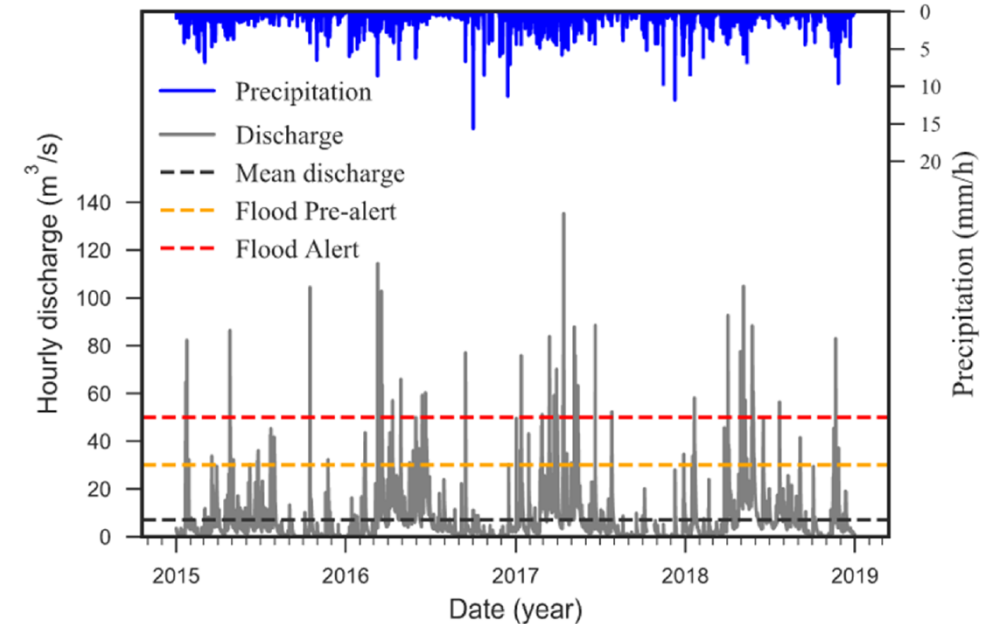
### ▼ CLASIFICACIÓN Valores categóricos



Article

### Flood Early Warning Systems Using Machine Learning Techniques: The Case of the Tomebamba Catchment at the Southern Andes of Ecuador

Paul Muñoz <sup>1,2,\*</sup>, Johanna Orellana-Alvear <sup>1,2</sup>, Jörg Bendix <sup>3</sup>, Jan Feyen <sup>4</sup> and Rolando Célleri <sup>1,2</sup>



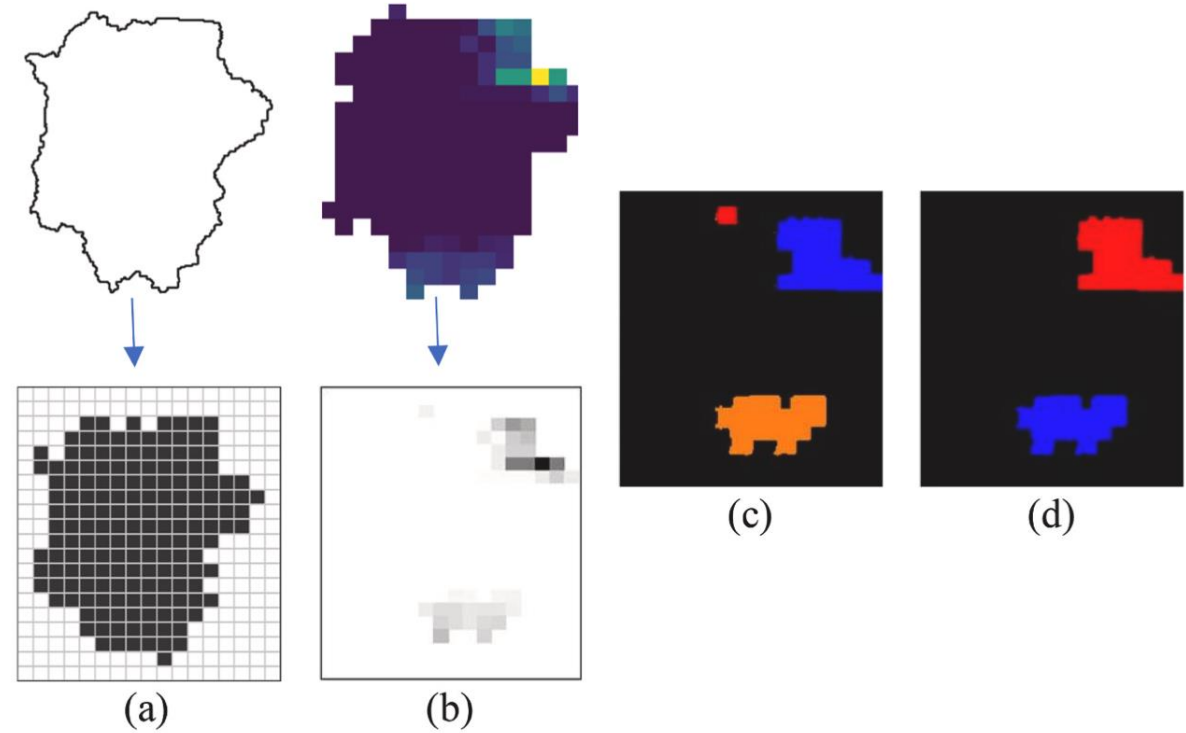
Alert  
Pre-alert  
No-alert



# TAREAS DE ML

## Aprendizaje no supervisado

- ▼ Clustering
- ▼ Predicción de conexiones
- ▼ Reducción de datos
- ▼ ...



Environmental Modelling and Software 160 (2023) 105582



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Environmental Modelling and Software

journal homepage: [www.elsevier.com/locate/envsoft](https://www.elsevier.com/locate/envsoft)



Near-real-time satellite precipitation data ingestion into peak runoff forecasting models

Paul Muñoz<sup>a,b,\*</sup>, Gerald Corzo<sup>c</sup>, Dimitri Solomatine<sup>c,d,e</sup>, Jan Feyen<sup>f</sup>, Rolando Céleri<sup>a,b</sup>



# PRINCIPALES FAMILIAS DE ALGORITMOS EN ML

## ▼ REGRESIVOS

*Regresión logística*

## ▼ ÁRBOLES DE DECISIÓN

*Bosque aleatorio*

## ▼ REDES NEURONALES

*Perceptrón multicapa*

## ▼ BASADO EN INSTANCIAS

*K vecinos más cercanos*

## ▼ BAYESIANOS

*Naïve Bayes*

# ¿Y LOS MODELOS TRADICIONALES BASADOS EN PROCESOS FÍSICOS?

- **El modelado basado en física es la opción ideal**

Siempre que la calidad/cantidad de datos lo permita...

- **Extrapolación limitada sin entrenamiento especializado**

Aumento de datos e ingeniería de características

- **Falta de interpretabilidad - ¿qué hay dentro de la caja?**

Las condiciones cambiantes pueden añadir ruido a los modelos

Represas, reglas operativas, fuentes de entrada



# ¿CUÁNDO NO USAR ML?

- ¿NO ESTÁ CLARO EL OBJETIVO?

La correlación no implica causalidad

- NO TIENES SUFICIENTES DATOS

Variabilidad espaciotemporal

- NO TIENES LOS DATOS ADECUADOS

Pronóstico de inundaciones repentinas usando datos diarios o más

- TIENES DEMASIADOS DATOS

Basura entra -> Basura sale



# EN LA PRÁCTICA

- LENGUAJES DE PROGRAMACIÓN
  - PYTHON

Incluso aquí varias librerías:

- **SCIKIT-LEARN (PRINCIPIANTES)**
- PYTORCH (DL POR FACEBOOK)
- TENSORFLOW (DL POR GOOGLE)
- KERAS

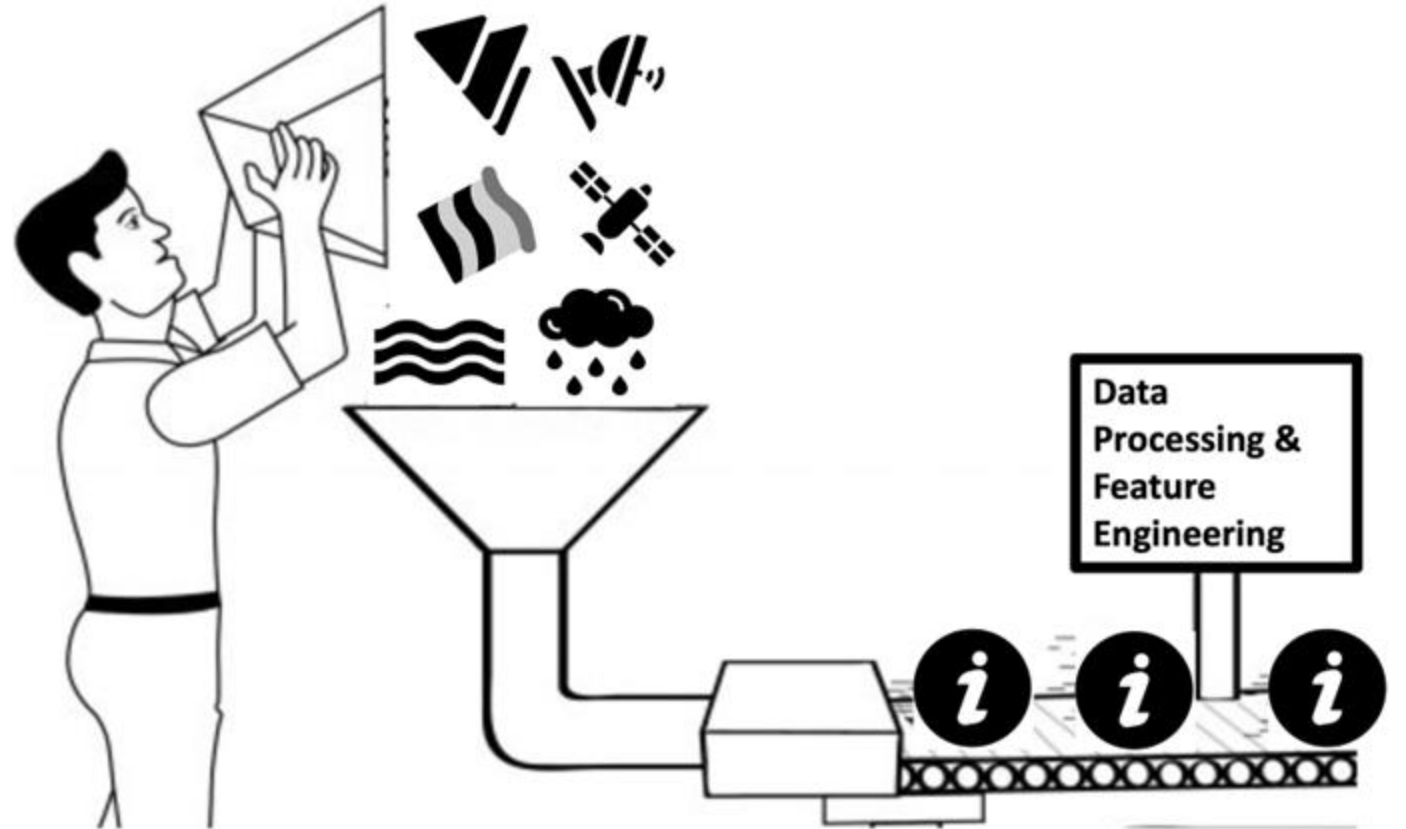
- R
- C++
- MATLAB





# PROCESAMIENTO DE DATOS

- Revisar CALIDAD/CANTIDAD
- Tratar con valores faltantes
- Valores atípicos, extremos
- Transformaciones  
Escalar/normalizar



# ALGO DE TERMINOLOGÍA

- CALIBRACIÓN → TRAINING  
→ VALIDATION
- VALIDACIÓN → TESTING
- SOBREAJUSTE
- INPUT DATA →  
CARACTERÍSTICAS/PREDICTORES  
UNA VARIABLE (PRECIPITACIÓN) PUEDE TENER  
VARIAS CARACTERÍSTICAS (ESTACIONES/PIXELES)