

Entrega 1: Descripción y Formulación del Objetivo

Clasificación de la Calidad del Agua del Río de la Plata

**Contexto y Relevancia:** El Río de la Plata, compartido por Argentina y Uruguay, constituye una fuente vital de agua dulce cuya calidad se ve influenciada por factores tanto naturales como Humanas. En este contexto, las técnicas de aprendizaje automático ofrecen herramientas avanzadas para analizar parámetros físico-químicos y microbiológicos, permitiendo predecir con mayor precisión la calidad del agua y generar modelos útiles para la gestión ambiental.

Los conjuntos de datos que se utilizarán en este proyecto cumplen con la normativa nacional vigente para el monitoreo de la calidad del agua en ríos y otros cuerpos superficiales, y se proponen como modelo base para su aplicación en otras regiones del país, como para los ríos de Tierra del Fuego.

**Objetivo:** Predecir la calidad del agua a partir del análisis de parámetros físico-químicos y microbiológicos registrados entre los años 2013 al 2024 y que servirá como modelo base para evaluar la calidad del agua de Tierra del Fuego y de otras regiones del país.

**Tipo de problema:** Este es un problema de **clasificación supervisada**, ya que la variable objetivo representa categorías ordinales del estado de la calidad del agua, como “levemente deteriorada, deteriorada, muy deteriorada y extremadamente deteriorada”.

Variable Posibles:

1. OD (Oxígeno disuelto)

2. pH

3. colif\_totales\_ufc\_100ml

4. escher\_coli\_ufc\_100ml

5. enteroc\_ufc\_100ml

6. Nitrato\_mg\_l

7. NH4\_mg\_l (Amonio)

8. P\_total\_l\_mg\_l (Fósforo total)

9. Fosf\_ortofos\_mg\_l (Fósforo ortofosfato)
10. DBO\_mg\_l (Demanda Biológica de Oxígeno)

11. DQO\_mg\_l (Demanda Química de Oxígeno)

12. Turbiedad\_NTU

13. Hidr\_Deriv\_Petr\_ug\_l (Hidrocarburos derivados del petróleo)

14. Cr\_total\_mg\_l (Cromo total)

15. Cd\_total\_mg\_l (Cadmio total)

16. Clorofila\_a\_ug\_l

17. Microcistina\_ug\_l

18. ICA (Índice de Calidad del Agua)

Modelos posibles por utilizar:

Red Neuronal Artificial (ANN)	Gradient Boosting (XGBoost, LightGBM, CatBoost)
Árboles de decisión	Máquinas de vectores de soporte (SVM)
Random Forest	k-Vecinos más Cercanos (k-NN)