

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.

**Tipo de problema:** Clasificación

**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.

Variable Posibles:

1. OD (Oxígeno disuelto)	10. DBO_mg_l (Demanda Biológica de Oxígeno)
2. pH	11. DQO_mg_l (Demanda Química de Oxígeno)
3. colif_totales_ufc_100ml	12. Turbiedad_NTU
4. escher_coli_ufc_100ml	13. Hidr_Deriv_Petr_ug_l (Hidrocarburos derivados del petróleo)
5. enteroc_ufc_100ml	14. Cr_total_mg_l (Cromo total)
6. Nitrato_mg_l	15. Cd_total_mg_l (Cadmio total)
7. NH4_mg_l (Amonio)	16. Clorofila_a_ug_l
8. P_total_l_mg_l (Fósforo total)	17. Microcistina_ug_l
9. Fosf_ortofos_mg_l (Fósforo ortofosfato)	18. ICA (Índice de Calidad del Agua)

**Fuente de datos:** Existen 12 data sets que fueron descargados desde la plataforma Kaggle y CiAM, específicamente de los siguientes enlaces. Se destaca que el mismo aun no fue analizado por la comunidad de la plataforma Kaggle.

<https://www.kaggle.com/datasets/palomachiacchiara/muestreos-de-calidad-de-agua-de-la-riiglo/data>

<https://ciam.ambiente.gob.ar/repositorio.php?tid=1&stid=105&did=408#>

Modelos Posibles:

Regresión logística	Gradient Boosting (XGBoost, LightGBM, CatBoost)
Árboles de decisión	Máquinas de vectores de soporte (SVM)
Random Forest	k-Vecinos más Cercanos (k-NN)