

# GeoAuPredict (GAP): Aprendizaje profundo para predicción geoespacial de oro

Proyecto final — Curso de Deep Learning

Edward Calderón y Equipo

13 de octubre de 2025

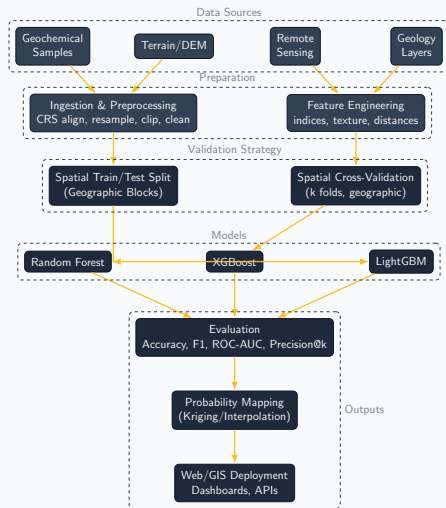
# Motivación y problema

- Relevancia: exploración, costos, riesgo.
- Enfoques clásicos: levantamientos, interpolación, geoestadística.
- Desafíos: datos heterogéneos, autocorrelación espacial, faltantes.
- Objetivo: aprender de datos multimodales para predecir ocurrencia de Au.

- Fuentes: DEM, índices de teledetección, capas geológicas, ensayos geoquímicos, distancias a estructuras.
- Unificación CRS, remuestreo, recorte del área de estudio, normalización.
- Particiones espaciales; balance de clases.

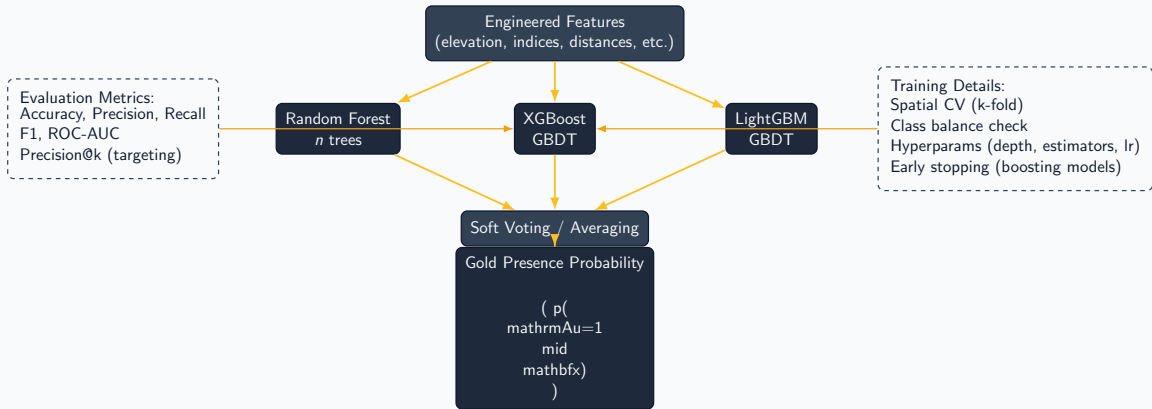
- Correlaciones y distribuciones; atípicos y faltantes.
- Mapas de ensayos y variables; clusters y hotspots.
- Ideas para ingeniería de características.

# Flujo de la tubería



- Ensamblados: Random Forest, XGBoost, LightGBM.
- Validación cruzada espacial (bloques geográficos).
- Mapas de probabilidad e incertidumbre.

# Arquitectura (Ensamble)



- Métricas: Exactitud, Precisión, Recall, F1, ROC-AUC, Precision@k.
- Detalles: early stopping (boosting), chequeos de hiperparámetros.
- Resultados clave: LightGBM AUC **0.9243**; Ensamble (producción) AUC **0.9208**; AUC por bloques espaciales **0.86**.



# Comparación con métodos clásicos

- Baselines: kriging, regresión, WoE, SVM (AUC típica  $\approx 0.82$ ).
- Mejora: +~**22%** AUC ( $0.82 \rightarrow 0.9208/0.9243$ ).
- Fortalezas: no linealidad y multimodalidad; debilidades clásicas: estacionariedad/linealidad.

- Dashboards web, exportación GIS, APIs.
- Guiar muestreo, priorizar zonas, soporte a decisión.
- Limitaciones: shift de dominio, disponibilidad de datos, generalización.

**¿Por qué DL vs geoestadística?**

**Evitar sobreajuste**

**Interpretabilidad**

**Transferencia**

**Incertidumbre**

**SOTA**

**Limitaciones**

**Operación**

Interacciones no lineales; multimodalidad; escala. CV espacial; regularización; early stopping; hold-outs no contiguos.

Importancia de variables, SHAP, modelos sustitutos.

Shift de dominio; fine-tuning; entrenamiento multi-región.

MC-dropout, ensambles, regresión cuantílica (trabajo futuro).

Citar TorchGeo/RS DL; comparar enfoque/métricas; novedad.

Datos escasos, ruido de sensores, shift, interpretabilidad.

Ingesta continua; inferencia; integración GIS; feedback del campo.

# Conclusiones

- Novedad: integración multimodal, CV espacial, mapas de probabilidad.
- Impacto: menores costos, mejor focalización, escalabilidad.
- Futuro: incertidumbre, adaptación de dominio, explicabilidad.

¡Gracias!

Preguntas bienvenidas.

► Demo

[github.com/edwardcalderon/GeoAuPredict](https://github.com/edwardcalderon/GeoAuPredict)