



Paso 3/3

¡Bien hecho! Tu proyecto ha sido aprobado. 🎉
Puede que el revisor haya sugerido alguna recomendación para mejorarlo, así que asegúrate de comprobarlo.

[Vista](#)



Descripción del proyecto

Los datos se almacenan en tres archivos:

- `gold_recovery_train.csv` — el dataset de entrenamiento **descargado**
- `gold_recovery_test.csv` — el dataset de prueba **descargado**
- `gold_recovery_full.csv` — el dataset fuente **descargado**

Los datos se indexan con la fecha y la hora de adquisición (`date`). Los parámetros cercanos en el tiempo suelen ser similares.

Algunos parámetros no están disponibles porque fueron medidos o calculados mucho más tarde. Por eso, algunas de las características que están presentes en el conjunto de entrenamiento pueden estar ausentes en el conjunto de prueba. El conjunto de prueba tampoco contiene objetivos.

El dataset fuente contiene los conjuntos de entrenamiento y prueba con todas las características.

Tienes a tu disposición los datos en bruto que solamente fueron descargados del almacén de datos. Antes de construir el modelo, comprueba que los datos sean correctos. Para ello, utiliza nuestras instrucciones.

Instrucciones del proyecto

1. Prepara los datos

1.1. Abre los archivos y examina los datos.

Ruta de acceso a los archivos:

- `/datasets/gold_recovery_train.csv`
- `/datasets/gold_recovery_test.csv`
- `/datasets/gold_recovery_full.csv`

1.2. Comprueba que el cálculo de la recuperación sea correcto. Calcula la recuperación de la característica `rougher.output.recovery` mediante el conjunto de entrenamiento. Encuentra el *EAM* entre tus cálculos y los valores de la característica. Facilita los resultados.

1.3. Analiza las características no disponibles en el conjunto de prueba. ¿Cuáles son estos parámetros? ¿Cuál es su tipo?

1.4. Realiza el preprocesamiento de datos.

2. Analiza los datos

2.1. Observa cómo cambia la concentración de metales (*Au*, *Ag*, *Pb*) en función de la etapa de purificación.

2.2. Compara las distribuciones del tamaño de las partículas de la alimentación en el conjunto de entrenamiento y en el conjunto de prueba. Si las distribuciones varían significativamente, la evaluación del modelo no será correcta.

2.3. Considera las concentraciones totales de todas las sustancias en las diferentes etapas: materia prima, concentrado rougher y concentrado final. ¿Observas algún valor anormal en la distribución total? Si es así, ¿merece la pena eliminar esos valores de ambas muestras? Describe los resultados y elimina las anomalías.

3. Construye el modelo

3.1. Escribe una función para calcular el valor final de *sMAPE*.

3.2. Entrena diferentes modelos. Evalúalos aplicando la validación cruzada. Elige el mejor modelo y pruébalo utilizando la muestra de prueba. Facilita los resultados.

Utiliza estas fórmulas para las métricas de evaluación:

$$\text{sMAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|) / 2} \times 100\%$$

$$\text{sMAPE final} = 25\% \times \text{sMAPE(rougher)} + 75\% \times \text{sMAPE(final)}$$

Evaluación del proyecto

Hemos definido los criterios de evaluación para el proyecto. Léelos con atención antes de pasar al ejercicio.

Esto es lo que los revisores buscarán cuando evalúen tu proyecto:

- ¿Has preparado y analizado los datos adecuadamente?
- ¿Qué modelos has desarrollado?
- ¿Cómo has comprobado la calidad del modelo?
- ¿Has seguido todos los pasos de las instrucciones?
- ¿Has respetado la estructura del proyecto y explicado los pasos realizados?
- ¿Cuáles son tus hallazgos?
- ¿Has mantenido el código limpio y has evitado su duplicación?

Ya tienes las hojas informativas y los resúmenes de los capítulos anteriores, así que ya puedes empezar.

¡Buena suerte!

¿Qué te ha parecido el proceso de revisión del código? Por favor, cuéntanoslo.
Esto nos ayudará a mejorar la experiencia educativa.

Share your thoughts