

Passo 3/3

Muito bem! Seu projeto foi aprovado.👏  
O revisor pode ter deixado algumas sugestões para que você possa tornar o projeto ainda melhor, então não se esqueça de verificá-las.  
[Visualizar](#)



## Descrição do projeto

Os dados são armazenados em três arquivos:

- `gold_recovery_train.csv` — [download](#) do conjunto de dados de treinamento
- `gold_recovery_test.csv` — [download](#) do conjunto de dados de teste
- `gold_recovery_full.csv` — [download](#) do conjunto de dados fonte

Os dados são indexados com a data e hora da aquisição (característica `data` ). Os parâmetros que estão próximos uns dos outros em termos de tempo geralmente são semelhantes.

Alguns parâmetros não estão disponíveis porque foram medidos e/ou calculados muito mais tarde. Por isso, algumas das características presentes no conjunto de treinamento podem estar ausentes do conjunto de teste. O conjunto de teste também não contém objetivos.

O conjunto de dados de origem contém os conjuntos de treinamento e teste com todas as características.

Você tem os dados brutos, recebidos diretamente do cliente. Antes de construir o modelo, verifique a exatidão dos dados. Para isso, use nossas instruções.

## Instruções do projeto

### 1. Prepare os dados

1.1. Abra os arquivos e estude os dados.

Caminho para os arquivos:

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

1.2. Verifique se a quantidade retirada foi calculada corretamente. Usando o conjunto de treinamento, calcule a quantidade retirada para a característica `rougher.output.recovery` . Encontre o *EAM* entre seus cálculos e os valores das características. Forneça resultados.

1.3. Analise as características não disponíveis no conjunto de teste. Quais são esses parâmetros? Qual é o tipo deles?

1.4. Faça o pré-processamento de dados.

### 2. Analise os dados

2.1. Observe como a concentração de metais (*Au*, *Ag*, *Pb*) muda dependendo do estágio de purificação.

2.2. Compare as distribuições de tamanho de partícula de minério no conjunto de treinamento e no conjunto de teste. Se as distribuições variarem significativamente, a avaliação do modelo estará incorreta.

2.3. Considere as concentrações totais de todas as substâncias em diferentes estágios: minério bruto, concentrado bruto e concentrado final. Você percebe algum valor anormal na distribuição total? Se você fizer isso, vale a pena remover esses valores de ambas as amostras? Descreva os resultados e elimine as anomalias.

### 3. Construir o modelo

3.1. Escreva uma função para calcular o valor final *sMAPE*.

3.2. Treine modelos diferentes. Avalie-os usando validação cruzada. Escolha o melhor modelo e teste-o usando a amostra de teste. Forneça resultados.

Use estas fórmulas para métricas de avaliação:

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

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

## Avaliação do projeto

Reunimos os critérios de avaliação do projeto. Leia com atenção antes de passar para o caso.

Veja o que os revisores analisarão ao revisar seu projeto:

- Você preparou e analisou os dados adequadamente?
- Que modelos você desenvolveu?
- Como você verificou a qualidade do modelo?
- Você seguiu todos os passos das instruções?
- Você seguiu a estrutura do projeto e explicou as etapas realizadas?
- Quais são as suas conclusões?
- Você manteve o código limpo e evitou a duplicação de código?

Você tem suas folhas de revisão e resumos dos capítulos anteriores, então você está pronto para ir em frente.

Boa sorte!

Avançar

O que você achou do processo de revisão de código? Por favor nos informe. Isso nos ajuda a melhorar a experiência educacional.

Share your thoughts