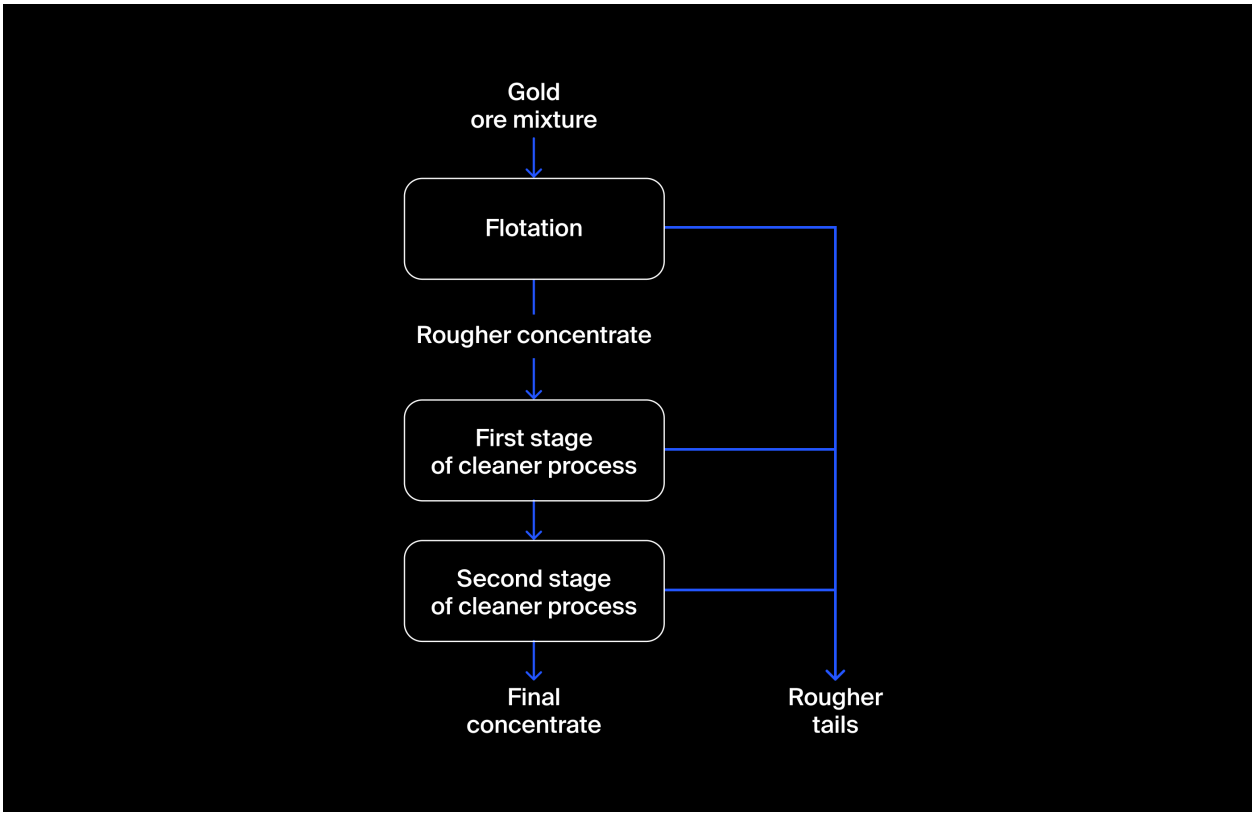
2. The Technological Process

> How is gold extracted from ore? Let's look at the stages of this process.

Mined ore undergoes primary processing to get the ore mixture or rougher feed, which is the raw material used for flotation (aka the rougher process). After flotation, the material goes through a two-stage purification.



Let's break down the process:

1. Flotation

The gold ore mixture is fed into the float banks to obtain a rougher Au concentrate and rougher tails (product residues with a low concentration of valuable metals).

The stability of this process is affected by the volatile and non-optimal physicochemical state of the flotation pulp (a mixture of solid particles and liquid).

2. Purification

The rougher concentrate undergoes two stages of purification. After purification, we have the final concentrate and new tails.

Data description

Technological process

- *Rougher feed* — raw material
- *Rougher additions* (or *reagent additions*) — flotation reagents: *Xanthate, Sulphate, Depressant*
  - *Xanthate* — promoter or flotation activator
  - *Sulphate* — sodium sulphide for this particular process
  - *Depressant* — sodium silicate
- *Rougher process* — flotation
- *Rougher tails* — product residues
- *Float banks* — flotation unit
- *Cleaner process* — purification
- *Rougher Au* — rougher gold concentrate
- *Final Au* — final gold concentrate

Stage parameters

- *air amount — volume of air*
- *fluid levels*
- *feed size* — feed particle size

- *feed rate*

Feature naming

Here's how you name the features:

`[stage].[parameter_type].[parameter_name]`

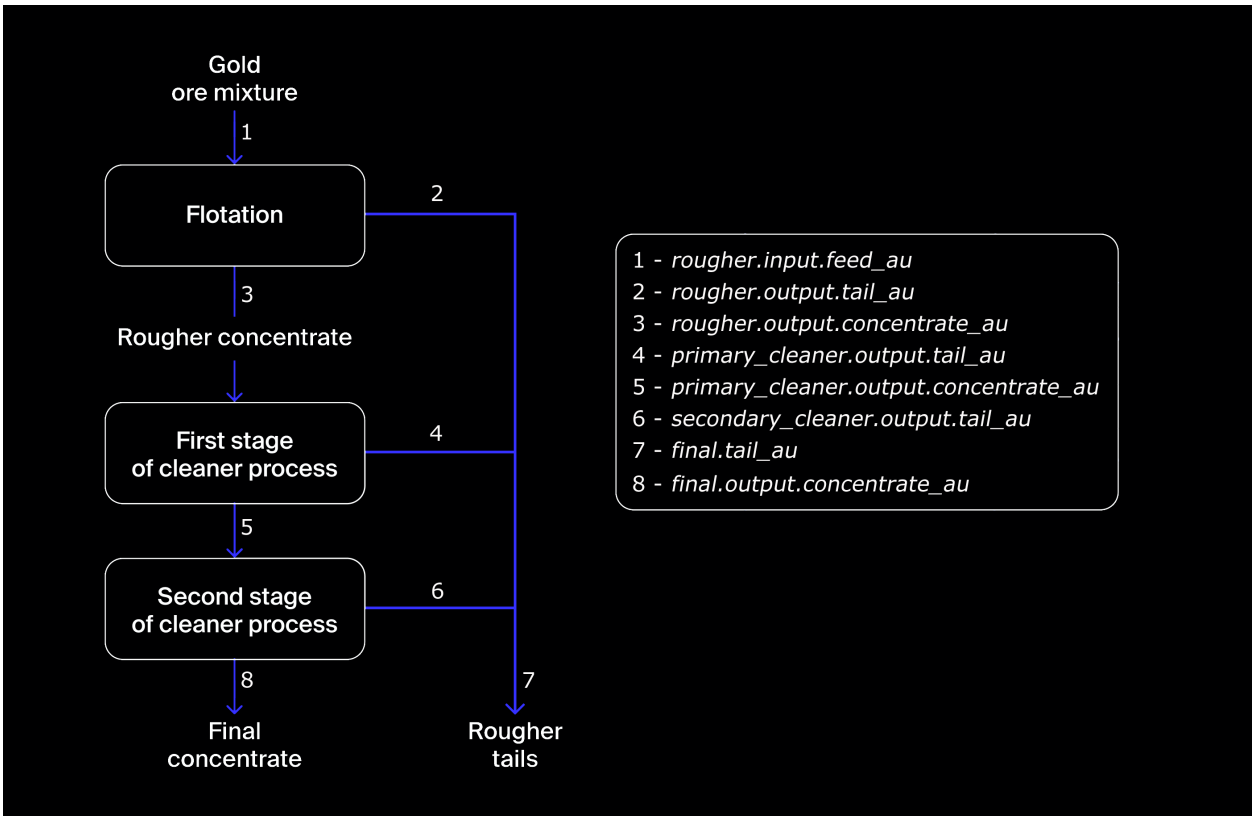
Example: `rougher.input.feed_ag`

Possible values for `[stage]`:

- *rougher* — flotation
- *primary_cleaner* — primary purification
- *secondary_cleaner* — secondary purification
- *final* — final characteristics

Possible values for `[parameter_type]`:

- *input* — raw material parameters
- *output* — product parameters
- *state* — parameters characterizing the current state of the stage
- *calculation* — calculation characteristics



Recovery calculation

Use the following formula to simulate the process of the recovery of gold from gold ore.

$$\text{Recovery} = \frac{\text{C} \times (\text{F} - \text{T})}{\text{F} \times (\text{C} - \text{T})} \times 100\%$$

- *C* — the share of gold in the concentrate right after flotation (for finding the rougher concentrate recovery)/after purification (for finding the final concentrate recovery)
- *F* — the share of gold in the feed before flotation (for finding the rougher concentrate recovery)/in the concentrate right after flotation (for finding the final concentrate recovery)
- *T* — the share of gold in the rougher tails right after flotation (for finding the rougher concentrate recovery)/after purification (for finding the final concentrate recovery)

To predict the coefficient, you'll need to find the share of gold in the concentrate and the tails. Note that both final and rougher concentrates matter.

Evaluation metric

To solve the problem, we will need a new metric. It is called **sMAPE**, or symmetric Mean Absolute Percentage Error.

It is similar to the MAE, but it's expressed in relative values instead of absolute ones. Why is it symmetrical? It takes into account the scale of both the target and the prediction.

Here's how *sMAPE* is calculated:

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

Definitions:

$$y_i$$

- The value of the target for the observation with the *i* index in the set used to measure quality.

$$\hat{y}_i$$

- The value of the prediction for the observation with the *i* index, for example, in the test set.

$$\text{N}$$

- The number of observations in the set.

$$\sum_{i=1}^{\text{N}}$$

- The sum of all observations in the set (***i*** takes values from 1 to *N*).

We need to predict two values:

1. Rougher concentrate recovery `rougher.output.recovery`
2. Final concentrate recovery `final.output.recovery`

The final metric includes the two values:

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