

# Predição de Qualidade no Processo de mineração

POC: Algoritmo de Predição de Silica

# Introdução



- A sílica é considerada uma **impureza**, e seu teor influencia diretamente a **qualidade do produto final**, os **custos operacionais** e o **impacto ambiental** da operação.
- O objetivo do modelo é prever o **percentual de sílica (% Silica Concentrate)** presente no concentrado de minério de ferro ao final do processo de flotação.



A previsão é realizada através de um modelo de regressão linear que considera:

- **Características do minério de entrada** (qualidade da polpa antes da flotação)
- **Variáveis de controle do processo** (vazão de ar, flotação e variáveis operacionais)
- **Histórico temporal do processo**, respeitando a natureza de série temporal dos dados



**Antecipação de problemas de qualidade:** Permite identificar tendências de aumento da sílica antes que o produto final

**Tomada de decisão em tempo quase real:** Engenheiros podem ajustar variáveis do processo (ar, nível, reagentes) com base na previsão

**Maior estabilidade do processo:** Reduz oscilações operacionais e variações indesejadas na qualidade do concentrado

## Base de Dados - Características do Minério

Essas colunas descrevem a composição e propriedades físicas do minério de ferro alimentado no processo de flotação

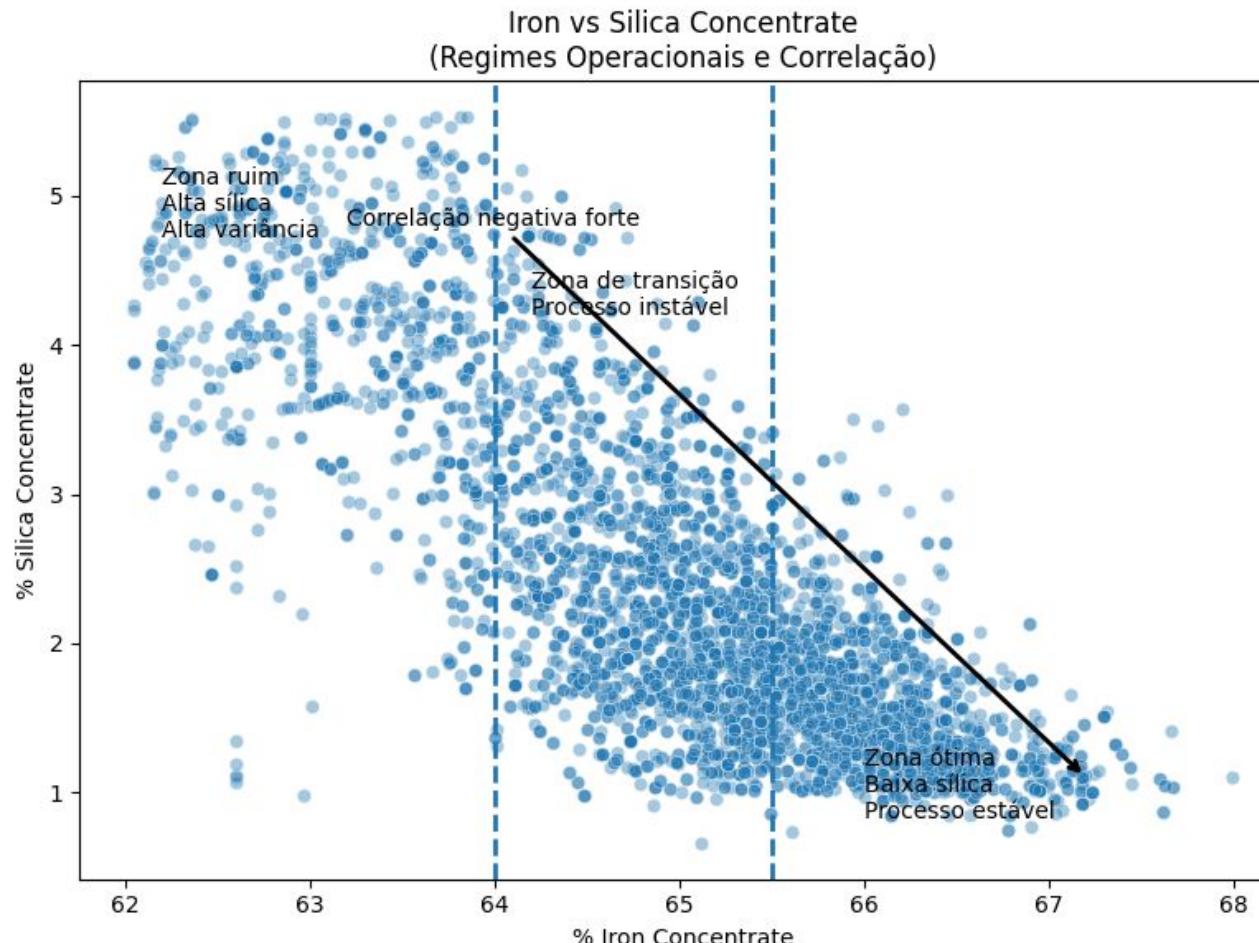
Colunas	Descrição Detalhada
<b>% Iron Feed</b>	Percentual de ferro (Fe) no minério bruto que entra nas células de flotação. Representa a qualidade inicial do feed.
<b>% Silica Feed</b>	Percentual de sílica ( $\text{SiO}_2$ ), principal impureza que precisa ser separada do ferro. Alto valor indica minério de baixa qualidade.
<b>% Iron Concentrate</b>	Percentual de ferro no concentrado final após flotação. Medida de qualidade do produto (quanto maior, melhor).
<b>% Silica Concentrate</b>	Percentual de sílica remanescente no concentrado final. É o target principal (quanto menor, melhor recuperação). Medida em laboratório.

# Base de Dados - Variáveis de Controle de Processo (

Essas são as variáveis operacionais ajustadas pelos operadores para otimizar a separação ferro/sílica nas colunas de flotação

Categoria	Coluna	Descrição Detalhada
Reagentes e Fluxos	starch (reagent) flow	Fluxo de amido, reagente que ajuda a separar sílica do ferro por adsorção seletiva.
	Amina Flow	Fluxo de amina, coletores que fazem a sílica flutuar para remoção.
	Ore Pulp Flow	Taxa de alimentação de polpa de minério (sólido + água) no circuito.
Propriedades da Polpa	Ore Pulp pH	Acidez da polpa (ideal ~9-11 para flotação de óxidos de ferro).
	Ore Pulp Density	Densidade da polpa (controla viscosidade e bolhas de ar).
Colunas de Flotação (7x)	Flotation Column XX Air Flow	Fluxo de ar injetado na coluna XX para formar bolhas que levam impurezas para cima.
	Flotation Column XX Level	Nível da polpa na coluna XX (controle PID para estabilidade do processo).

# Análise exploratória



## Relação fortemente negativa

Quanto **maior o Iron, menor a Sílica**.

- Nuvem inclinada da esquerda-cima → direita-baixo
- Isso indica **correlação negativa forte**

## Forma de “leque” (heterocedasticidade)

- O processo é **instável em regimes ruins e estável em regimes bons** (Em Iron baixo (~62–64) → Sílica varia MUITO (1 até 5+))
- Em Iron alto (~66–68) → Sílica fica **bem concentrada** (~1–1.5) (O processo é **instável em regimes ruins e estável em regimes bons**)

## Por que Silica é o target certo

- Iron **satura** rápido
- Sílica **explode quando algo dá errado**
- Prever Sílica = prever **perda de eficiência**

# Fluxograma



# Resultados

## Modelo com Top 10 features por importance

- Quase o mesmo desempenho do Full Data
- Menor risco de overfitting
- Mais fácil de explicar e manter

Experimento	Features usadas	R <sup>2</sup>	RMSE	Desvio Padrão do Alvo	RMSE / Std	Ganho vs Baseline	Avaliação Geral
Remoção de colunas	Subconjunto reduzido	0.181	0.9308	1.0286	0.90	9.54%	<span style="color:red;">✗</span> Fraco / Marginal
Full Data	Todas as features	0.654	0.6054	1.0286	0.59	41.17%	<span style="color:green;">✓</span> Bom
Feature Importance (Top 10)	10 features mais relevantes	0.647	0.6111	1.0286	0.59	40.61%	<span style="color:green;">✓</span> Bom

# Conclusão e Próximos Passos

## Conclusões

1. **Seleção de features baseada em importância é eficaz**, mantendo performance com menor complexidade
2. **Modelo atual é estável e confiável**, com split temporal adequado
3. **Há espaço claro para melhoria**, principalmente via modelagem não-linear e features temporais

## Próximos Passos Recomendados

1. **Testar modelos não-lineares**
  - Random Forest, XGBoost, LightGBM
  - Validação com TimeSeriesSplit
2. **Engenharia de features temporais**
  - Lags do alvo
  - Médias e desvios móveis
3. **Benchmark com baseline forte**
  - Naive forecast
  - Média móvel