









# Introdução à linguagem de programação Python: conceitos básicos e requisitos para aplicação de algoritmos de IA

EDITAL FAPERGS 06/2024 - PROGRAMA DE PESQUISA E DESENVOLVIMENTO VOLTADO A DESASTRES CLIMÁTICOS

Leandro Rodrigues Oviedo Engenheiro Químico Doutor em Nanociências – UFN



GRUPO DE PESQUISA
EM NANOMATERIAIS APLICADOS

## Tópicos abordados

Conceitos gerais – O que é Python? Funcionalidades do Python Ambientes: execução local e execução em nuvem Principais bibliotecas (pesquisa científica – análise exploratória e algoritmos de IA (construção de dataframes, regressão não linear, teste de normalidade, correlação...) Estudos de casos

## O que é Python?

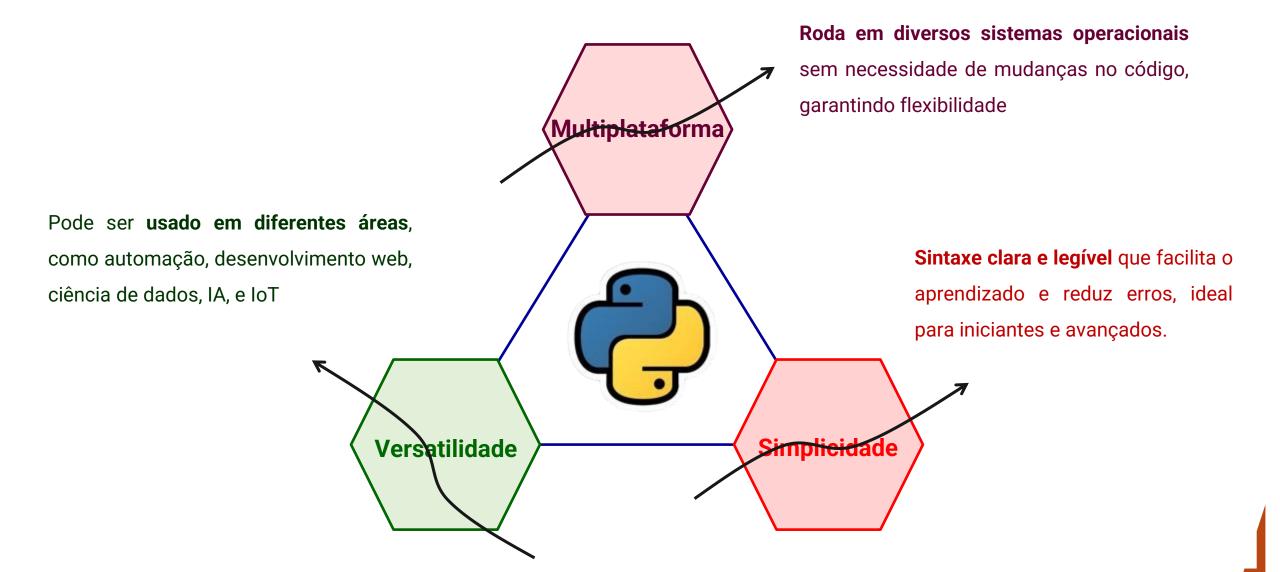


Linguagem de programação de uso geral, que permite criar uma grande variedade de aplicações diferentes. Destaca-se pela sua versatilidade e facilidade de uso, tornando-se uma das linguagens mais comuns atualmente.

Linguagem amplamente utilizada em desenvolvimento de software, análise de dados, inteligência artificial, automação de tarefas e criação de aplicativos web, contando com vasta biblioteca padrão e uma comunidade ativa que facilita encontrar soluções para diversos projetos.

Python é uma linguagem interpretada, ou seja, não necessita de compilação prévia, o que acelera o desenvolvimento.

### Python e suas vantagens



#### **Funcionalidades**

#### Operações matemáticas e análise estatística

- Regressão linear e não linear;
- Análise exploratória dos dados (EDA)
- Teste de normalidade, análise de correlação e teste de hipóteses.

#### Tratamento e visualização de dados

- Plotagem de gráficos interativos e não interativos;
- Construção de dataframes, normalização de dados e visualizações (scatterplot, violinplot, histogramas, gráficos 3D...);

#### Construção e compilação de algoritmos avançados

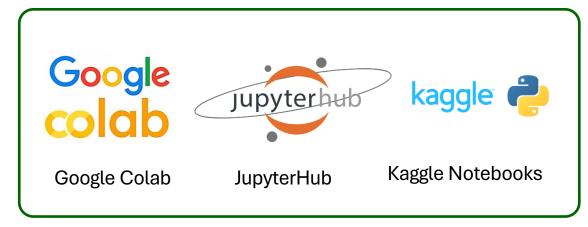
- Algoritmos de inteligência artificial (IA);
- Análise de extração de features (feature extraction/selection);
- Planejamento de experimentos (design of experiments DOE);
- Elaboração de WebApp, páginas html e interfaces gráficas, dashboards e outros.

## Ambientes de execução

#### **Execução local** / Computador pessoal/IDE/editor



#### **Execução em nuvem** / Browser, servidores remotos

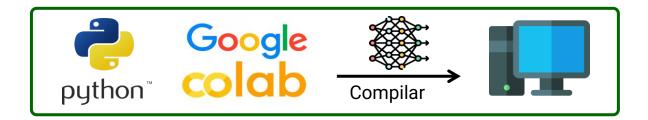


#### Principais bibliotecas

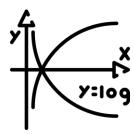
Versão +atual do software: Python 3.12.11

Bibliotecas +utilizadas na pesquisa científica-----

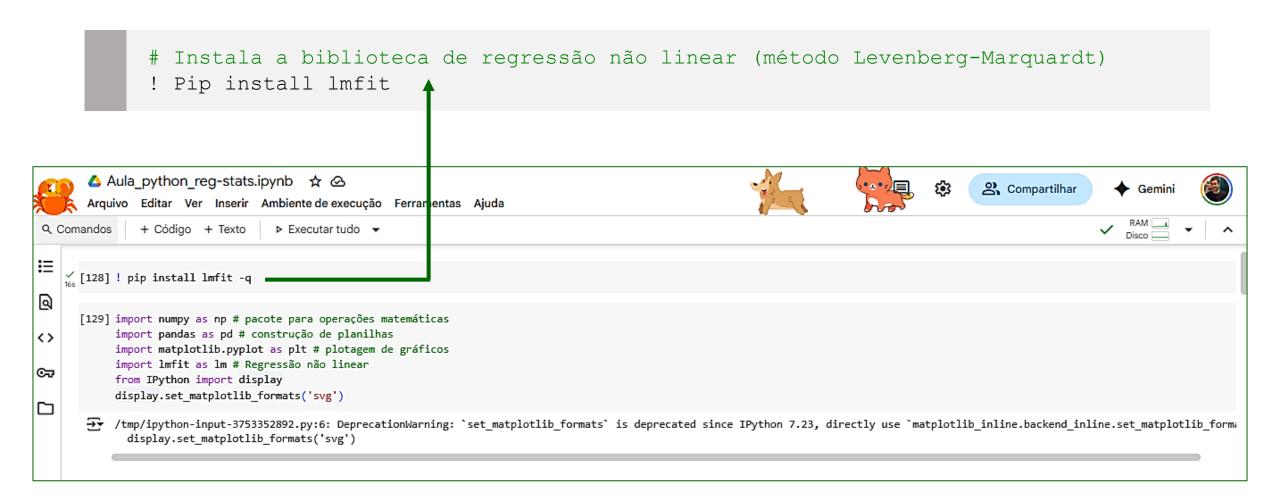
numpy # computação numérica e manipulação eficiente de arrays
pandas # manipulação e análise de dados tabulares
matplotlib / seaborn / plotly #visualização de dados, estática e interativa
scipy # matemática aplicada, otimização e ferramentas científicas
sympy # álgebra computacional e cálculo simbólico
statsmodels # análise estatística e modelagem econométrica
scikit-learn # machine learning e pré-processamento de dados
Levenberg-Marquardt # algoritmo para otimização e ajuste de modelos não lineares
tensorflow / keras # desenvolvimento e treinamento de redes neurais (multicamadas e profundas)

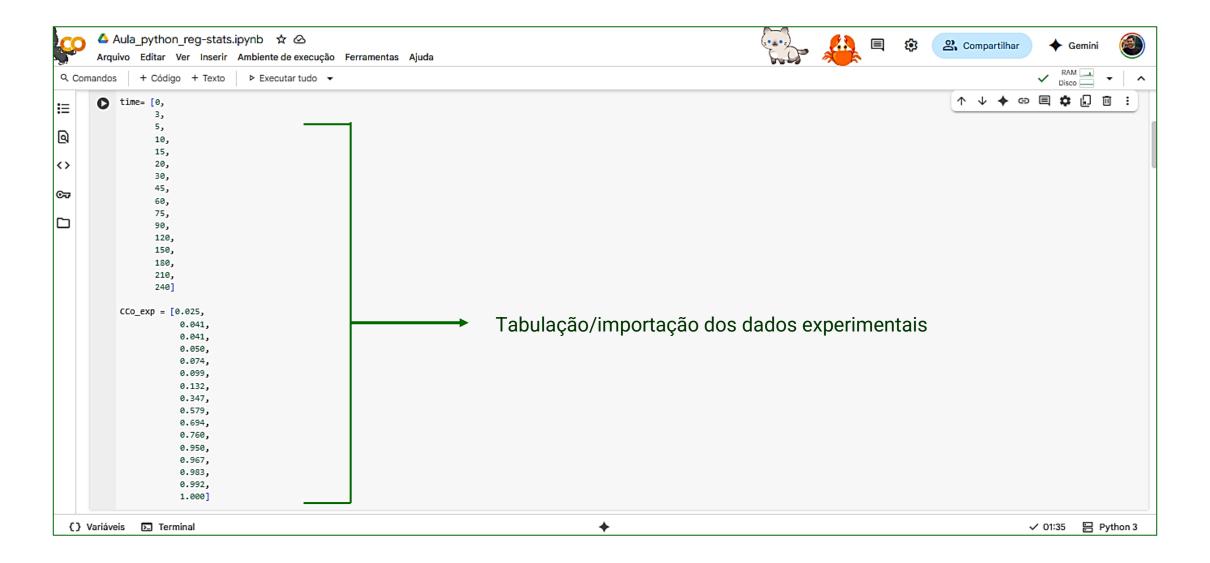


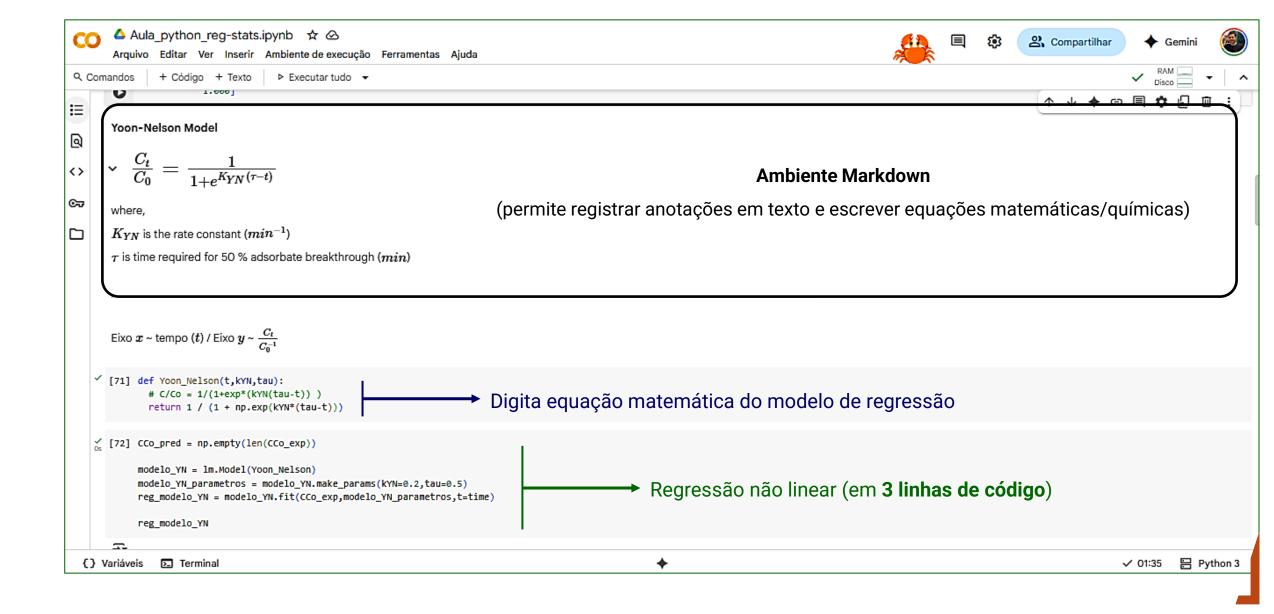
**Temática:** Deseja-se realizar uma regressão não linear dos dados experimentais de adsorção, utilizando um modelo cinético de 3 parâmetros. Para isso, defina qual sua equação, os parâmetros envolvidos e utilize as funções da biblioteca Levenberg-Marquardt. Qual a **confiabilidade** no modelo de regressão utilizado? Obtenha um relatório completo sobre **métricas** do modelo de regressão e compare com as métricas do modelo de 2 parâmetros. **Qual o tempo de processamento do algoritmo registrado em sua máquina?** 

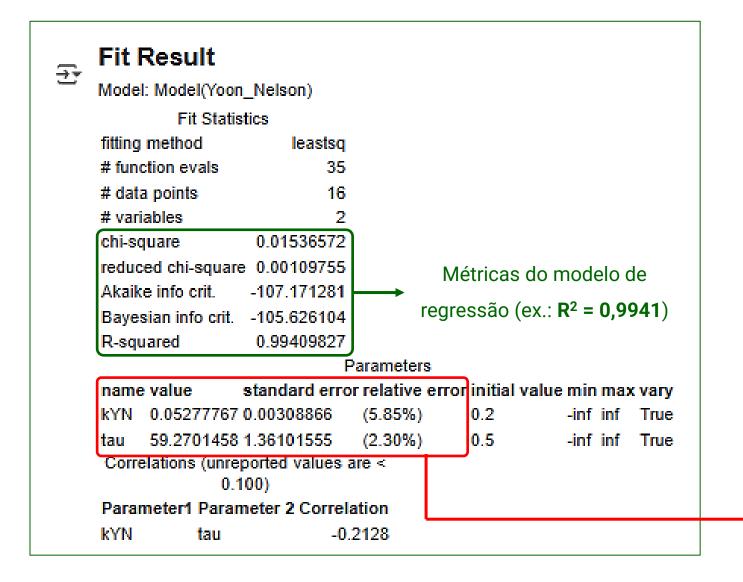


O método de **Levenberg-Marquardt** é um algoritmo usado para ajustar parâmetros de modelos não lineares por meio de otimização iterativa, combinando os métodos de Gauss-Newton e gradiente descendente para melhorar a convergência. Uma das bibliotecas que permitem sua implementação é lmfit (>! pip install lmfit > import lmfit as lm)









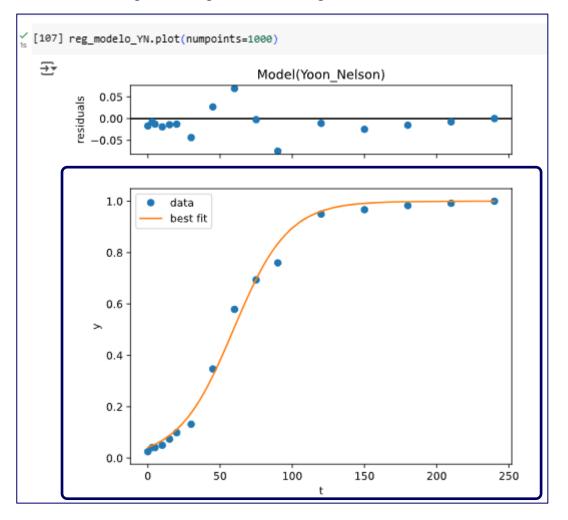
A função **Imfit** nos fornece um relatório completo de regressão

Estimativa dos parâmetros da equação (neste caso, temos **2 parâmetros**)

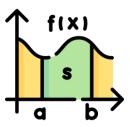
#### Cálculo do coeficiente de determinação ajustado (R2<sub>adj</sub>)

```
[73] K = len(reg_modelo_YN.params)
     N = len(time)
     print(f'Tenho {K} parâmetros')
     print(f'Tenho {N} dados',N)
     print('Tenho {0:d} parâmetros e {1:d} dados'.format(K,N))
    Tenho 2 parâmetros
     Tenho 16 dados 16
     Tenho 2 parâmetros e 16 dados
       R<sup>2</sup> = sample R-square
        p = Number of predictors
        N = Total sample size.
[74] R2 = reg_modelo_YN.rsquared
     R2_ajd = 1-((1-R2)*(N-1)/(N-K-1))
     print(f'R2: {R2}\nR2 ajustado: {R2_ajd}')
     R2: 0.9940982658793455
     R2 ajustado: 0.9931903067838601
```

#### Plotagem do gráfico de regressão não linear



**Temática:** Deseja-se determinar a distribuição do tamanho de partícula de uma nanopartícula de óxido de cobre (CuO-NPs) a partir de imagens de microscopia eletrônica de varredura de alta resolução (MEV-FEG). Para isso, um código em Python pode ser construído para ler a imagem de MEV-FEF (formato .tif), detectar o número de partículas na micrografia, fornecer uma planilha com os valores da distribuição, plotar um histograma e printar os valores de média e desvio padrão para o tamanho/diâmetro de partícula. Para isso, será utilizado o YOLO (You Only Look Once), um algoritmo de detecção de objetos preciso e rápido, ideal para identificar e contar partículas em imagens de MEV/TEM/MEV-FEG.

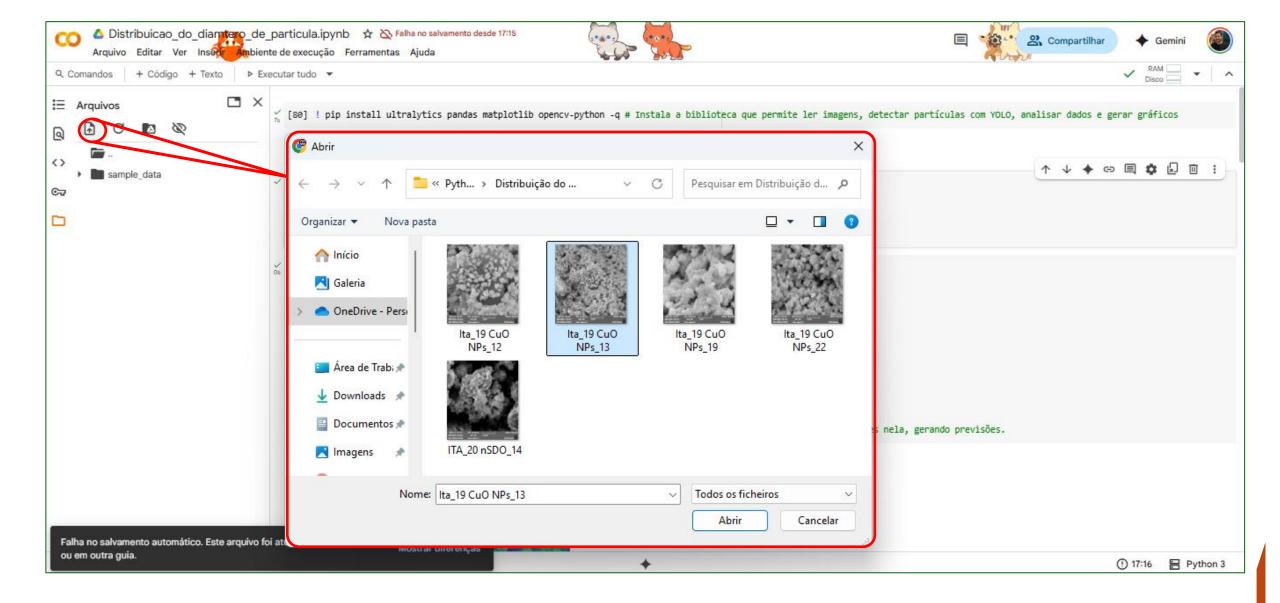


o algoritmo **YOLO** funciona processando toda a imagem de uma vez **e localizando múltiplos objetos simultaneamente**, retornando as coordenadas e dimensões de cada partícula detectada, o que facilita a extração de tamanhos e análise estatística da distribuição dos diâmetros.



Ferramenta que permite o carregando de arquivos no Python

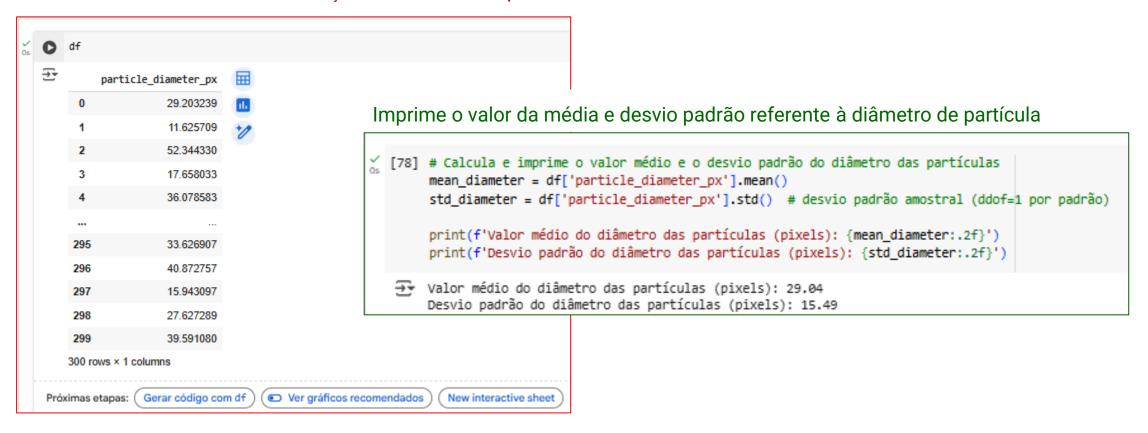
(Opcionalmente, pode-se carregar, ler e processar dados diretamente contidos no Google Drive)

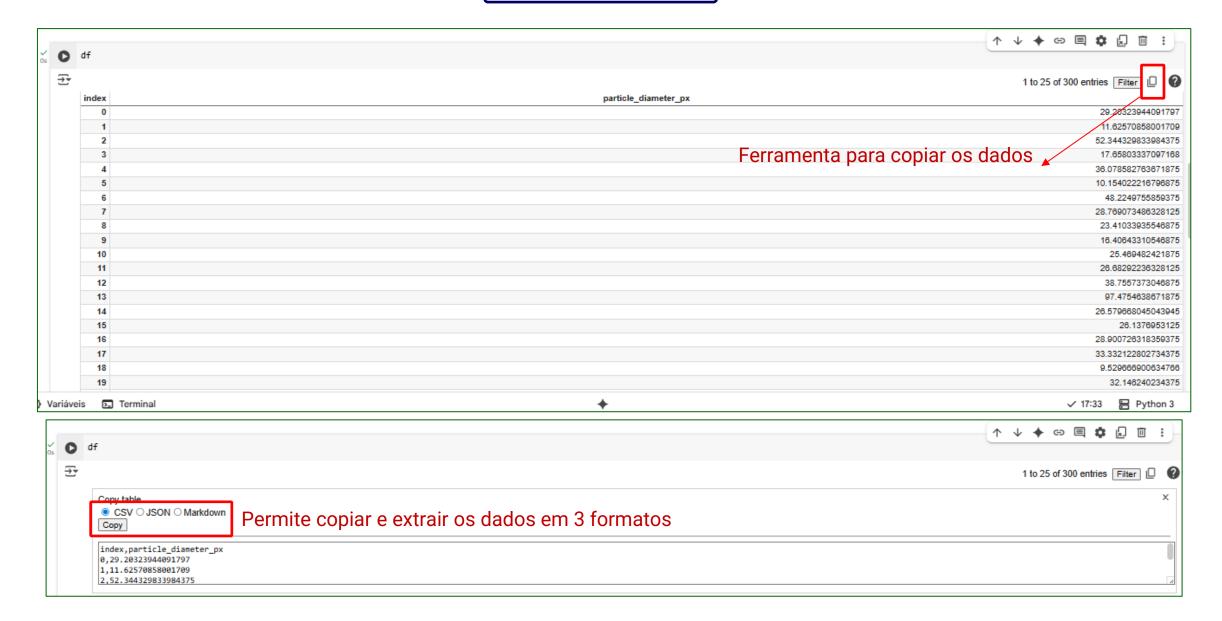


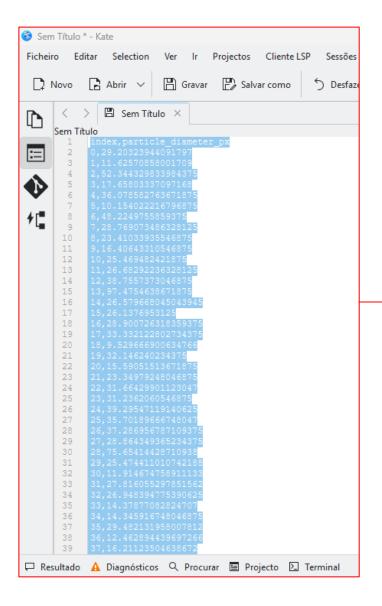




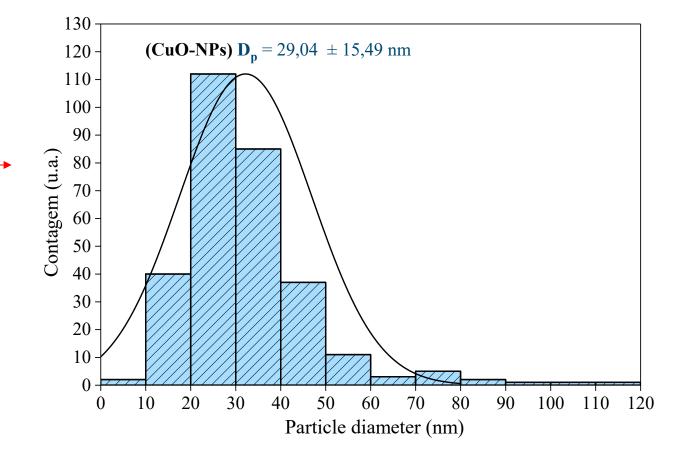
#### Fornece dataframe com a distribuição do diâmetro de partícula



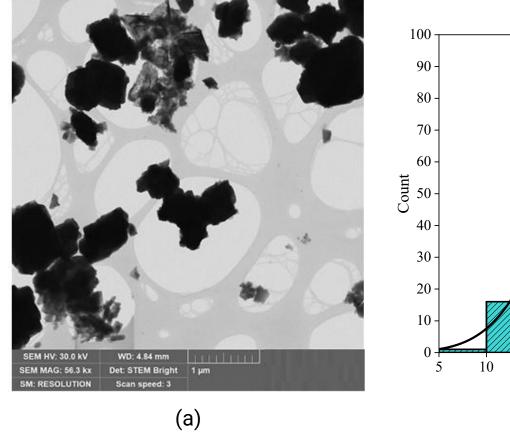


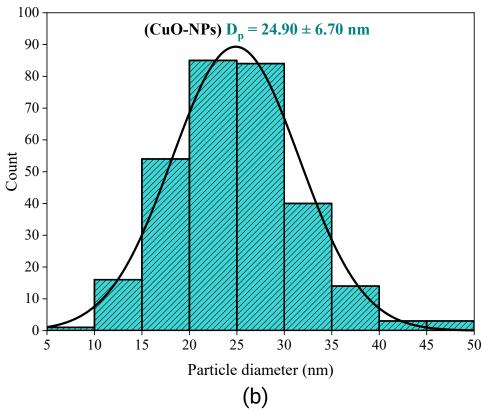


**Figura 2.** Distribuição do diâmetro médio de partícula das nSOD@CuO-NPs gerado a partir dos do Python e reproduzido em *software* OriginLab 2025b.

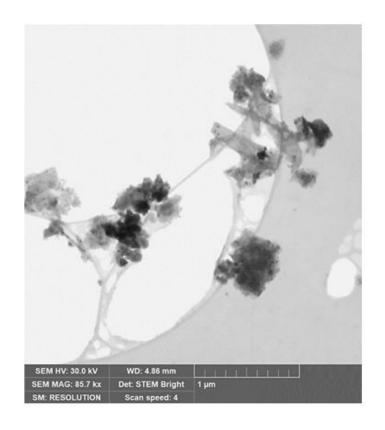


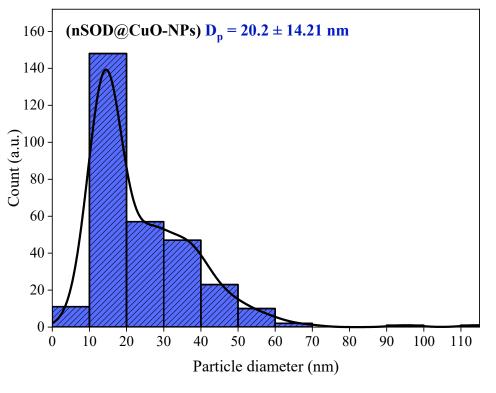
**Figura 3.** (a) Micrografias de microscopia eletrônica de transmissão (TEM) de CuO-NPs com ampliação de 56,3 kx e (b) distribuição do tamanho médio de partícula das CuO-NPs





**Figura 4.** (a) Micrografias de microscopia eletrônica de transmissão (TEM) de nSOD@CuO-NPs com ampliação de 85,7 kx e (b) distribuição do tamanho médio de partícula das nSOD@CuO-NPs





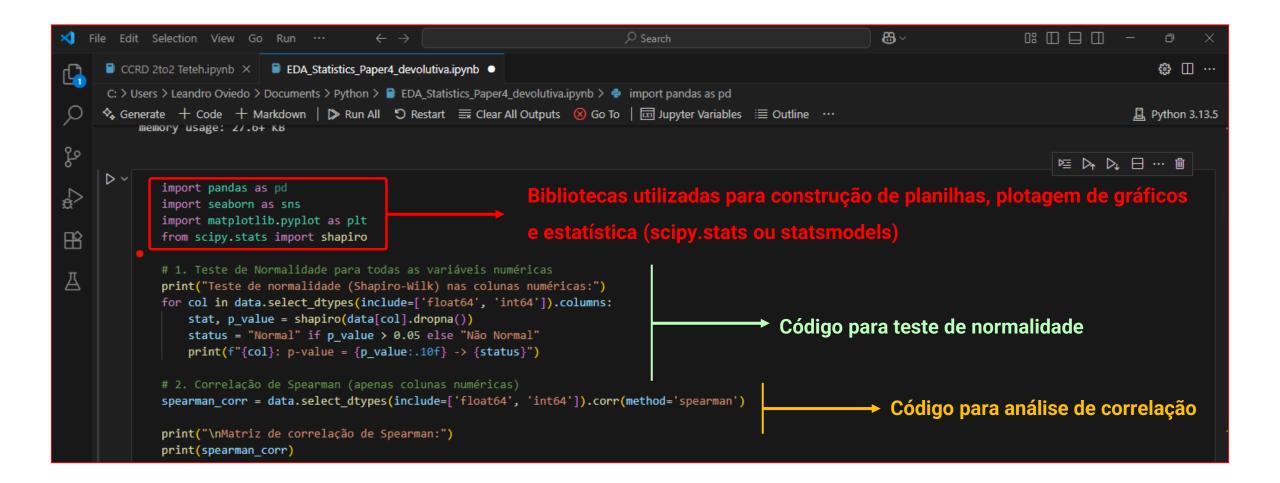
(a)

(b)

**Temática:** Deseja-se realizar uma Análise Exploratória dos Dados (EDA – Exploratory Data Analysis) para verificar qual a melhor condição de um tratamento avançado de água (fotocatálise heterogênea para a degradação de uma mistura binária de corantes sintéticos). Para isso, sugere-se realizar uma análise de correlação e visualização de dados para identificar a condição ótima do processo oxidativo.



**EDA** é o processo inicial de examinar e visualizar dados para entender suas características principais, detectar **padrões**, **outliers** e **relações** entre variáveis. Sugere-se realizar a **análise de correlação** antes de aplicar **machine learning supervisionado**, pois isso permite identificar dependências e redundâncias entre variáveis, auxiliando na seleção das features mais relevantes e evitando modelos superajustados (**overfitting**) ou enviesados.



```
print("\nMatriz de correlação de Spearman:")
print(spearman_corr)
# 3. Heatmap da correlação com paleta Greens e barra de escala (colorbar)
plt.figure(figsize=(7, 7))
sns.heatmap(
    spearman corr,
   annot=True,
                                                                                      Código para plotar a matriz de correlação
   cmap='Greens',
   cbar=True,
                                                                                      (biblioteca seaborn para gráficos)
   vmin=-1, vmax=1,
   square=True,
   linewidths=0.5,
   linecolor='gray'
                                                                                      Pode salvar em formatos png, svg, jpeg, PDF e tif.
plt.title('Matriz de Correlação de Spearman (Paleta Greens)')
plt.show()
```



Para salvar a Figura, pode-se utilizar a função plt.savefig('nome\_da\_figura.svg', dpi =600)

plt.tight\_layout() para ajustar automaticamente os espaços entre os elementos do gráfico para que não fiquem sobrepostos ou cortados,

Quanto maior o dpi, maior a qualidade de imagem (Python trabalha com **geração de Imagem** na faixa de **72 a 1400 dpi**)

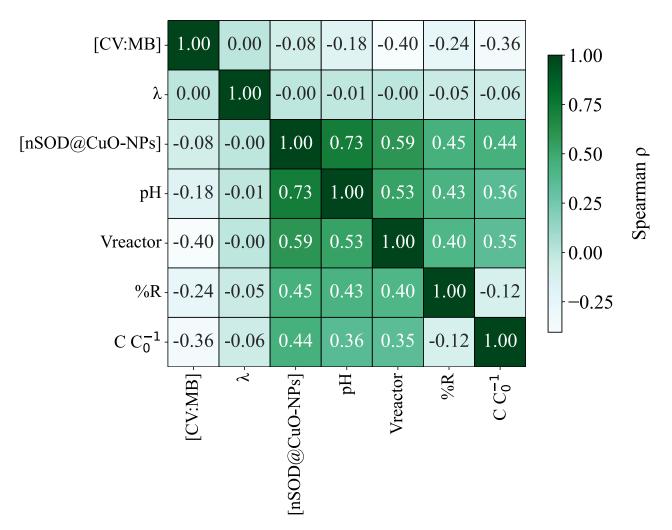


Figure 6. Spearman correlation for experimental data / Fonte: https://doi.org/10.1016/j.jwpe.2025.108508

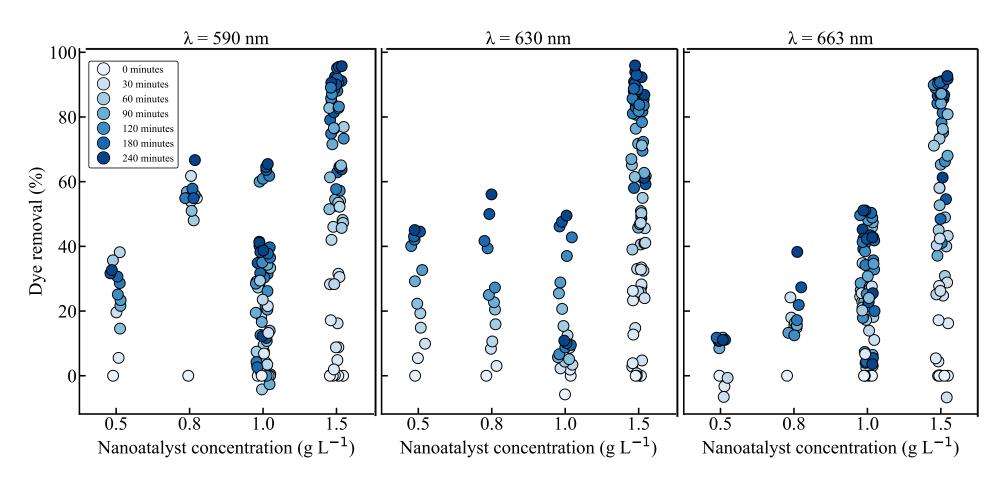


Figure 6. Effect of the nanocatalyst concentration and reaction time on dye removal (Sample size (N) for nanocatalyst concentration: 0.5 g L<sup>-1</sup> (N = 51), 0.8 g L<sup>-1</sup> (N = 51), 1.0 g L<sup>-1</sup> (N = 122), 1.5 g L<sup>-1</sup> (N = 176) /  $V_{reactor}$  = 100 mL | T = 25 ± 2°C |  $y_{CV}$  = 0.44 and  $y_{MB}$  = 0.56 | under visible radiation with 600 W m<sup>-2</sup>).

Fonte: <a href="https://doi.org/10.1016/j.jwpe.2025.108508">https://doi.org/10.1016/j.jwpe.2025.108508</a>

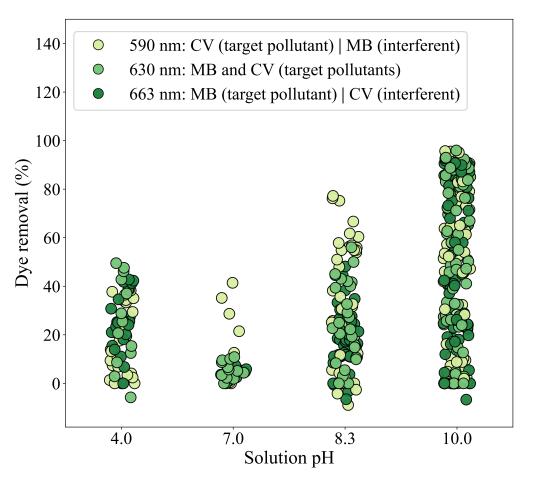
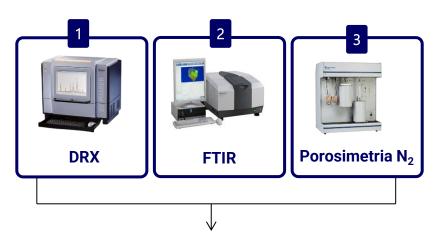
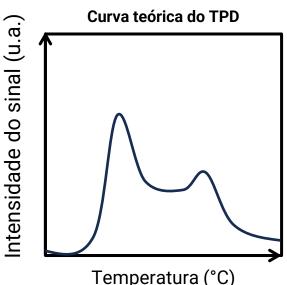


Figure 6. Effect of the pH and wavelength on dye removal. (Sample size (N): pH 4 (N = 35), pH 7 (N = 53), pH 8.3 (N = 136), pH 10 (N = 176)  $V_{reactor} = 100 \text{ mL} \mid T = 25 \pm 2^{\circ}\text{C} \mid y_{CV} = 0.44 \text{ and } y_{MB} = 0.56 \mid \text{under visible radiation with } 600 \text{ W m}_{-2}$ 

Fonte: https://doi.org/10.1016/j.jwpe.2025.108508

#### Estudos de casos



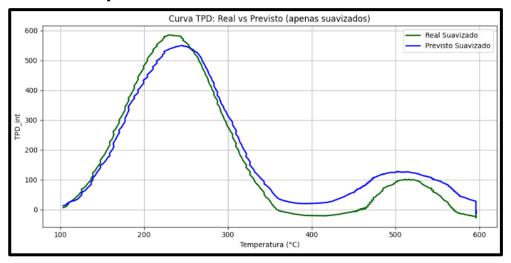


#### Estudo de Caso 4

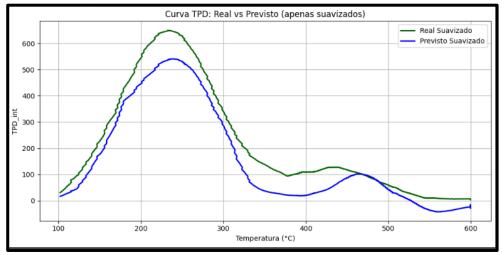
**Temática:** Há disponível um difratômetro de raios X (DRX), um espectrofotômetro no infravermelho por Transformada Fourier (FTIR) e um analisador de área (porosimetria de  $N_2$ ). Quero complementar os resultados dessas técnicas de caracterização com a técnica de Temperatura de Dessorção Programada (TPD) para determinar a acidez de um nanoadsorvente. No entanto, não há recursos financeiros nem disponibilidade do equipamento físico de TPD. Propõe-se a utilização de IA, com as técnicas acima e dados de TPD (retirados da literatura) para prever a curva TPD de um nanoadsorvente de interesse, cujos resultados de DRX, FTIR e Porosimetria de  $N_2$  estão disponíveis.

É possível **prever a curva TPD** para esse nanoadsorvente, tendo o resultado de caracterização das demais técnicas?

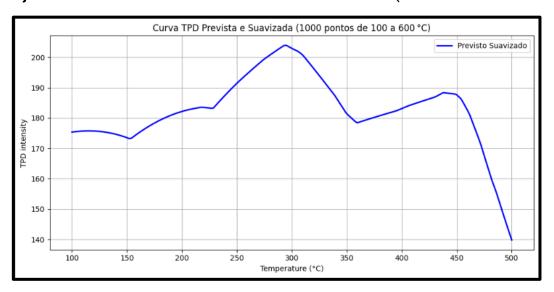
**Predição 1.** Nanozeólita FAU contendo 1% de Ni



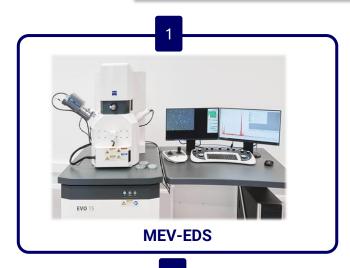
Predição 2. Nanozeólita FAU contendo 10% de Ni



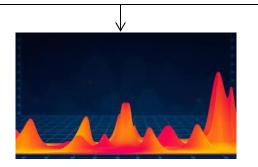
Predição 3. Nanozeólita FAU contendo 10% de Nb (amostra de interesse)



#### Estudos de casos







#### **Estudo de Caso 5**

**Temática:** Deseja-se prever a distribuição da acidez de nanomateriais (nanozeólitas modificadas e não modificadas com Nb) a partir de dados de adsorção em batelada (piridina), imagens de microscopia eletrônica de varredura e composição elementar (espectroscopia de energia dispersiva).

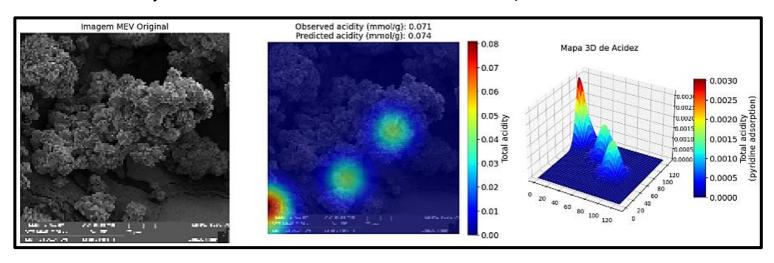
#### Substituição isomórfica

Átomos de Al e Si são trocados por átomos de Nb na estrutura zeolítica

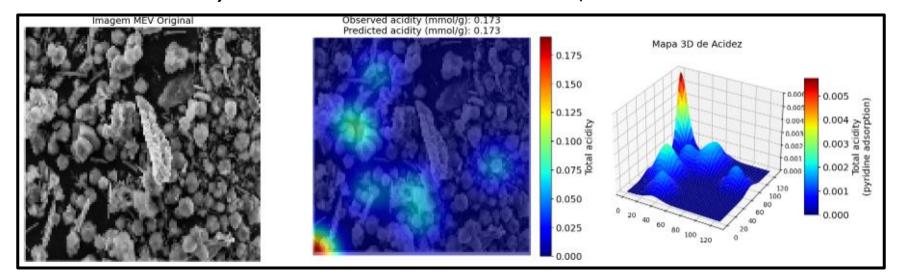


É possível **gerar um mapa de calor** para esse nanoadsorvente, mostrando a distribuição de uma alguma propriedade (acidez total) sobre sua estrutura

Predição 1. Nanozeólita FAU contendo 10% Nb a partir de MEV-FEG.



Predição 2. Nanozeólita FAU contendo 10% Nb a partir de MEV.



#### Particularidades

#### A linguagem de programação Python é...



- Case-sensitive, isto é, uma variável k difere de uma variável K);
- Zero-indexed, isto é, inicia a contagem de elementos em 0 (0, 1, 2, 3...).

#### Informações adicionais



- Pode ser processada com CPU local ou GPU (processamento da Google)
- Alguns códigos podem ser muito pesados e demorados, necessitando máquinas mais potentes ou softwares mais sofisticados para uma aplicação;
- Execução em nuvem é recomendada para projetos pequenos;
- Busque registrar suas linhas de códigos para manter/garantir originalidade.

#### Links úteis

#### **GitHub**



Modelos de caixa branca e cinza

https://github.com/LeandroOviedo

#### Boas práticas de programação



YouTube

https://www.youtube.com/watch?v=MwoA-pzzAQw





## **VAMOS PRATICAR???**



## **GRATO PELA ATENÇÃO!**





















# Introdução à linguagem de programação Python: conceitos básicos e requisitos para aplicação de algoritmos de IA

EDITAL FAPERGS 06/2024 - PROGRAMA DE PESQUISA E DESENVOLVIMENTO VOLTADO A DESASTRES CLIMÁTICOS

Leandro Rodrigues Oviedo Engenheiro Químico Doutor em Nanociências – UFN



GRUPO DE PESQUISA
EM NANOMATERIAIS APLICADOS