

**INSTITUTO INFNET**

ESCOLA SUPERIOR DE TECNOLOGIA

GRADUAÇÃO EM CIÊNCIA DE DADOS



Projeto de Bloco: Ciência de Dados Aplicada

**TP 3**

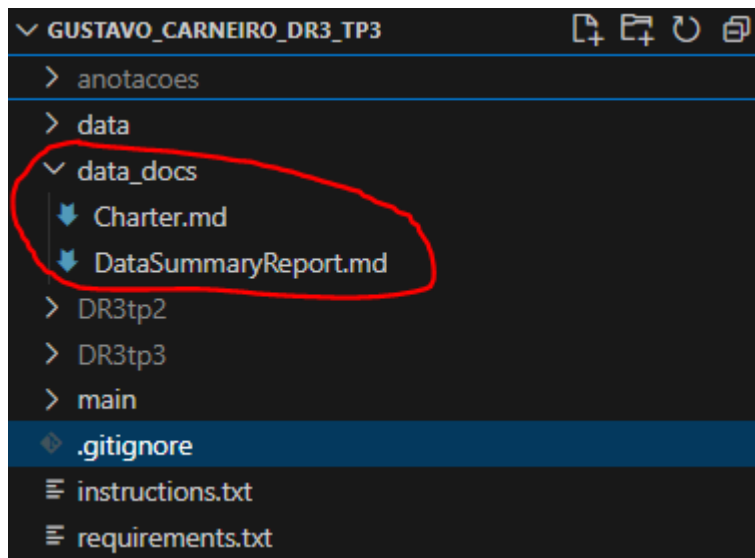
Alunos: Gustavo Carneiro Alves.

2024

Link: [https://github.com/GustavoAlvesInfnet/gustavo\\_carneiro\\_DR3\\_TP2](https://github.com/GustavoAlvesInfnet/gustavo_carneiro_DR3_TP2)

### 1. Revisão e Atualização da Documentação:

- Revise o **Project Charter** e o **Data Summary Report**, atualizando a documentação para refletir as novas funcionalidades e decisões tomadas nesta fase do projeto.
- Reavalie o problema de negócio à luz das novas ferramentas (como FastAPI e Selenium) e ajuste suas metas, se necessário.
- Atualize a descrição das fontes de dados utilizadas, considerando possíveis novas fontes obtidas com scraping dinâmico.



### 2. Criação de uma Aplicação com Múltiplas Páginas:

- Evolua a interface da sua aplicação em **Streamlit**, implementando **múltiplas páginas** e um **menu de navegação** que permita ao usuário transitar facilmente entre diferentes seções da aplicação.
- Cada página deve representar uma funcionalidade ou análise diferente, como a visualização de dados, gráficos interativos, upload/download de arquivos ou estatísticas geradas a partir dos dados coletados.

The screenshot shows a web application with a dark theme. On the left, there is a sidebar with a dropdown menu titled 'Selecione uma opção' containing the following items: 'Introdução' (highlighted), 'Introdução', 'Análise dinâmica', 'Upload para análise individual', and 'Análise Web'. The main content area is titled 'Introdução' and features the heading 'AI Learn - Trilha de Aprendizagem Personalizada'. Below the heading, it states: 'Um dos problemas do sistema de ensino atual é a dificuldade para criar uma trilha personalizada para cada aluno.' It then describes the application's purpose: 'Minha aplicação irá criar uma trilha individual baseada nas dificuldades dos alunos que servirá como auxílio extra para os estudos visando ajudar individualmente nos pontos fracos de cada um.' At the bottom, it provides a link to the data source: 'Link dos dados usados: [Kaggle](#)'.

### 3. Extração de Dados de Páginas Dinâmicas (Web Scraping):

- Utilize o **Selenium** para realizar o web scraping de páginas dinâmicas, se necessário. Caso seu projeto utilize uma fonte de dados que exija interação com elementos dinâmicos (como formulários ou carregamentos assíncronos), o Selenium será essencial.
- **Observação:** Se não houver necessidade de utilizar Selenium, concentre-se no aprimoramento dos dados coletados com BeautifulSoup ou APIs, mantendo a simplicidade quando possível.
- Armazene os dados obtidos em arquivos CSV ou TXT, organizando-os no diretório de **data/** para uso na aplicação.

Obs.: Estou usando o firefox para o scrapping com selenium

The screenshot shows a web browser displaying the Indeed job search results page. The search criteria are 'Design Editorial' and 'São Paulo'. The page shows a list of job results, with the first one being 'Social Media - SANTO ANDRÉ'. The job details include: 'PRO' status, location 'Santo André, SP', salary range 'R\$ 1.800 - R\$ 2.000 por mês', and a 'Candidatar-se' button. The page also features a sidebar with filters like 'Data do anúncio', 'Remoto', 'Salário', 'Tipo de vaga', 'Nível de escolaridade', 'Setor', 'Local', and 'Idioma da vaga'. The top navigation bar includes links for 'Página inicial', 'Avaliações de Empresas', 'Cargos e salários', 'Acessar', and 'Empresas/Anunciar vagas'.

Selecione uma opção

Analise Web

## Scrapping

Selecione o cargo e a região:

Cargo

Região

Quantidade de paginas

Executar o scraping

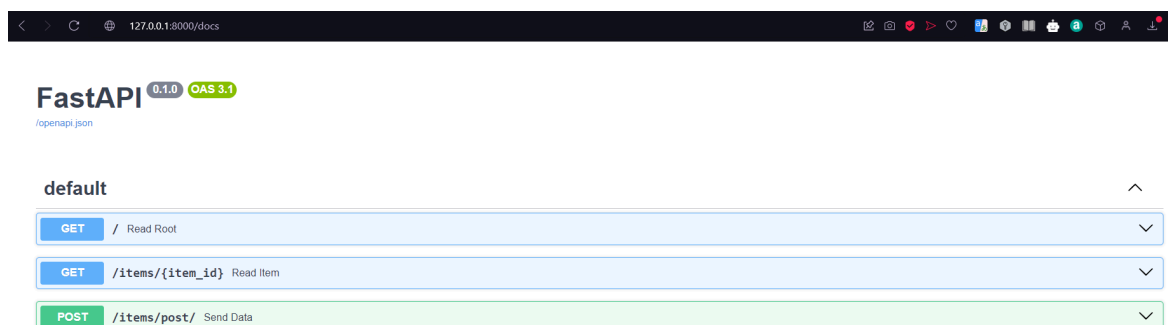
	title	company	rating
0	Arquiteto de Dados Pleno	Grupo Fleury	4,2
1	Analista de Desenvolvimento Pleno	BR-IT	Nao inform
2	Analista de Parcerias Sênior - BizDev	Serasa Experian	3,6
3	Analista de Dados	Beatriz Recrutamento	Nao inform
4	Analista de Dados Junior	Seedz	Nao inform
5	Analista de Dados	Rossana Rios	Nao inform
6	Analista de Suporte Banco de Dados (Oracle)	ALLTAK INDÚSTRIA E COMÉRCIO DE ADESIVOS LTDA	Nao inform
7	Analista de Backoffice Júnior	Amerinode	Nao inform
8	Analista de Programação Logístico (LOTS GR	Scanla Latin America	Nao inform
9	Analista de Sistema Sênior - Experiência na á	Millenium Brasil	Nao inform

#### 4. Desenvolvimento de APIs com FastAPI:

- Configure o ambiente de desenvolvimento para a criação de APIs com **FastAPI**.
- Crie uma **API simples** com rotas e endpoints que permitam interagir com os dados da aplicação.  
Exemplo de funcionalidades da API:
  - Consulta de dados (GET)
  - Envio de novos dados (POST)
- Implemente ao menos duas rotas em sua API, e documente-as adequadamente, explicando sua função e como elas podem ser usadas para interação com os dados.

```
uvicorn api:app --reload
```

Execute o comando acima dentro da pasta main



No parameters

Request body required application/json

```
{
  "texts": "Teste para o TP3"
}
```

Servers

These operation-level options override the global server options.

/

Execute Clear

Respostas

```

#transforma o json em dicionário
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS GITLENS COMMENTS
● PS D:\Infnet\2-semester\VSCode Projects\gustavo_carneiro_DR3_TP3> cd .\main\
○ PS D:\Infnet\2-semester\VSCode Projects\gustavo_carneiro_DR3_TP3\main> uvicorn api:app --reload
INFO: Will watch for changes in these directories: ['D:\\Infnet\\2-semester\\VSCode Projects\\gustavo_carneiro_DR3_TP3\\main']
INFO: Uvicorn running on http://127.0.0.1:8000 (Press CTRL+C to quit)
INFO: Started reloader process [28828] using StatReload
INFO: Started server process [27964]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: 127.0.0.1:61462 - "GET / HTTP/1.1" 200 OK
INFO: 127.0.0.1:61462 - "GET /favicon.ico HTTP/1.1" 404 Not Found
INFO: 127.0.0.1:61463 - "GET /docs HTTP/1.1" 200 OK
INFO: 127.0.0.1:61463 - "GET /openapi.json HTTP/1.1" 200 OK
Teste para o TP3
INFO: 127.0.0.1:61783 - "POST /items/post/ HTTP/1.1" 200 OK

```

## 5. Preparação para Uso de Inteligência Artificial com LLMs:

- Nesta etapa, comece a pensar nos dados que você coletou até agora e como eles podem ser utilizados em **tarefas baseadas em LLMs** nas próximas entregas.
- Considere os tipos de dados disponíveis e as possíveis aplicações com **LLMs**, como:
  - **Análise de Texto Gerado:** Usar os dados coletados para gerar resumos automáticos de textos longos, facilitando a compreensão e análise de documentos.
  - **Classificação de Sentimentos:** Aplicar um modelo de LLM para classificar os sentimentos em textos (positivos, negativos ou neutros) coletados de notícias, redes sociais, ou outras fontes.
  - **Perguntas e Respostas (Q&A):** Usar um LLM para construir um sistema de perguntas e respostas a partir dos dados disponíveis, respondendo a perguntas relevantes com base nos conteúdos coletados.
  - **Geração de Texto:** Automatizar a criação de relatórios ou insights com base nos dados brutos, utilizando LLMs para gerar textos descritivos.