

# VALIDAÇÃO DE DADOS PARA CONTRATOS

4º Edição

**DATA  
THON'S**  
2022

## **DATA THONS 4ª EDIÇÃO**

Acelerando a validação e assinatura de contratos CCT através de técnicas de Recuperação de Informação e Inteligência Artificial.





DATA DRIVEN  
INDÚSTRIA 4.0  
GESTÃO DE DOCUMENTOS  
BIG DATA

## \_GARIMPANDO OS DADOS

Poderosos **insights**, **oportunidades** de mercado, melhoria de processos e aumento da **eficiência** em uma empresa podem ser alcançados a partir da exploração inteligente dos seus dados.

Como?

. Desbloqueando os dados.

# **\_PROBLEMA**

A ONS, como garantidor da conformidade regulatória dos contratos, precisa validar todas as informações dos contratos CCTs\* como interveniente.


Desafios:

- Existem vários modelos de CCTs, variando entre as Transmissoras, com cláusulas diferentes. Tudo é validado manualmente.

**\*CONTRATO DE CONEXÃO AS INSTALAÇÕES DE TRANSMISSÃO**

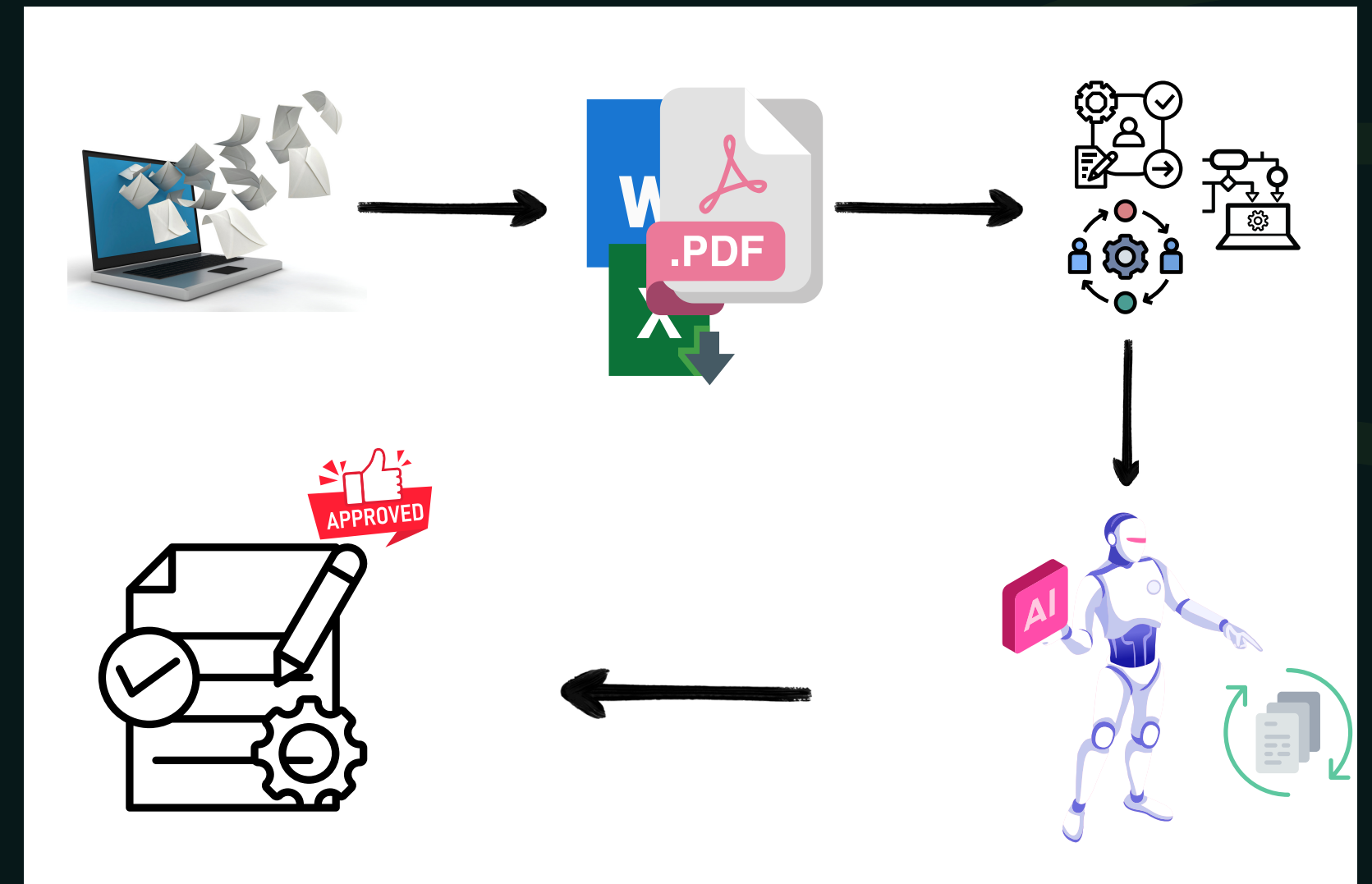


# \_O QUE VAMOS RESOLVER

- Extração e validação de informações do dados das instituições envolvidas
  - Extração e validação das informações de outorga
    - Contrato de concessão da transmissora
  - Extração e validação das informações do Parecer de Acesso
  - Qualificação dos diretores
  - Definição de Procedimentos de Rede
  - Validação das Cláusulas
    - Submissão aos Procedimentos de Rede e à Regulamentação da ANEEL
- 

# \_COMO RESOLVEMOS

Nosso algoritmo cruza as informações essenciais para a validação do CCT com outras bases de dados, afim de apontar se estão aptos a serem avaliados por um agente ou devem retornar para o remetente para correção de inconsistências.



# \_TECNOLOGIAS UTILIZADAS

Técnicas de recuperação de informação como extração de texto de documentos PDF, detecção de padrões e lógica fuzzy.

Linguagem e ferramentas:

- Python

Principais Bibliotecas:

- fuzzywuzzy
- pdfminer
- difflib.SequenceMatcher
- re (regex)



# \_METODOLOGIA

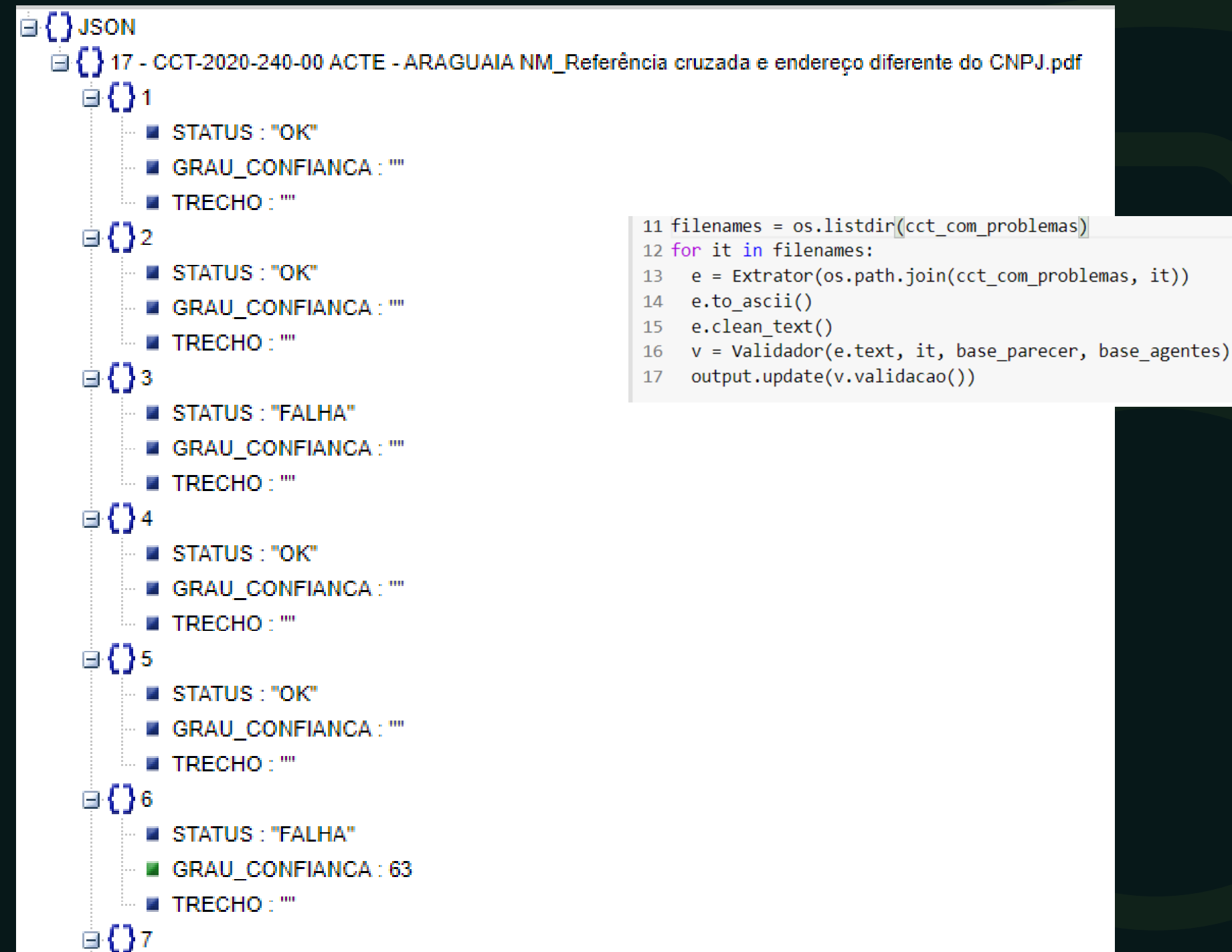
- O fluxo da nossa solução consiste em:
  - Extrair o texto de cada PDF
  - Fazer limpeza e padronização do texto.
  - Fazer as validações para cada item em seu respectivo método (classe Validador)

```
11 filenames = os.listdir(cct_com_problemas)
12 for it in filenames:
13     e = Extrator(os.path.join(cct_com_problemas, it))
14     e.to_ascii()
15     e.clean_text()
16     v = Validador(e.text, it, base_parecer, base_agentes)
17     output.update(v.validacao())
```

# \_METODOLOGIA

O script itera todos os PDFs presentes em um diretório, valida um por vez, e retorna os resultados para um dicionário.

Em seguida o dicionário é convertido em Json e pode ser visualizado como ao lado.



```
JSON
17 - CCT-2020-240-00 ACTE - ARAGUAIA NM_Referência cruzada e endereço diferente do CNPJ.pdf
1
  STATUS : "OK"
  GRAU_CONFIANCA : ""
  TRECHO : ""
2
  STATUS : "OK"
  GRAU_CONFIANCA : ""
  TRECHO : ""
3
  STATUS : "FALHA"
  GRAU_CONFIANCA : ""
  TRECHO : ""
4
  STATUS : "OK"
  GRAU_CONFIANCA : ""
  TRECHO : ""
5
  STATUS : "OK"
  GRAU_CONFIANCA : ""
  TRECHO : ""
6
  STATUS : "FALHA"
  GRAU_CONFIANCA : 63
  TRECHO : ""
7
  STATUS : "FALHA"
  GRAU_CONFIANCA : 63
  TRECHO : ""
```

```
11 filenames = os.listdir(cct_com_problemas)
12 for it in filenames:
13     e = Extrator(os.path.join(cct_com_problemas, it))
14     e.to_ascii()
15     e.clean_text()
16     v = Validador(e.text, it, base_parecer, base_agentes)
17     output.update(v.validacao())
```





## DUARTE JUNIOR

Eng de Energias | Cientista de Dados

[linkedin.com/in/duarte-junior-9530595a/](https://www.linkedin.com/in/duarte-junior-9530595a/)



## FLÁVIA LOPES

Cientista de Dados

[linkedin.com/in/lopesflavia](https://www.linkedin.com/in/lopesflavia)



## RANIERI IMPERATORI

Cientista de Dados

[linkedin.com/in/ranieri-imperatori/](https://www.linkedin.com/in/ranieri-imperatori/)



## ROBERTO ASANO

Especialista de Energias |  
Cientista de Dados

[linkedin.com/in/roberto-asano-253514a/](https://www.linkedin.com/in/roberto-asano-253514a/)