

Projeto de Pesquisa e Desenvolvimento

Sistema de Agentes Inteligentes para Automação da Classificação Fiscal em Setores Econômicos Brasileiros

Fernanda Ebina (1), Marlon Mendes (2), Paulo Victor A. Silva (3),
Victor Silva (5).

(1) ebina.fe@gmail.com (2) marlon.maxuel@educacao.mg.gov.br (3)
viktor.alkan1992@gmail.com (5) victorh.csilva20@gmail.com

29 de Outubro de 2025

Resumo

As empresas brasileiras enfrentam desafios na classificação fiscal de produtos e operações devido à complexidade da legislação tributária nacional. Por vezes podem haver erros na tributação que podem resultar em impactos financeiros e riscos de autuações. Este projeto propõe o desenvolvimento de um protótipo de sistema de agentes inteligentes para automatizar processos de classificação fiscal e cálculo de tributos. Utilizando Processamento de Linguagem Natural (*Natural Language Processing* - PLN) para interpretar descrições de produtos em Notas Fiscais Eletrônicas (NF-e) e um Motor de Regras de Negócio (*Business Rules Engine* - BRE). O sistema visa fornecer classificações e cálculos confiáveis para auxílio na validação dos impostos: ICMS; PIS e COFINS e IPI. A metodologia inclui a extração de conhecimento tributário de documentos legais com o auxílio de Grandes Modelos de Linguagem (*Large Language Models* - LLMs), espera-se com este projeto aumentar a eficiência operacional, reduzir erros de conformidade e criar uma base tecnológica para o desenvolvimento de soluções *TaxTech* (tecnologia para a área fiscal) no Brasil.

Palavras-chave: Automação Fiscal, Inteligência Artificial, Processamento de Linguagem Natural, Classificação Fiscal, ICMS, PIS, COFINS, IPI, *TaxTech*.

1 Introdução

O sistema tributário brasileiro é reconhecido por sua alta complexidade e por um volume de alterações normativas que impõe desafios à conformidade das empresas (CASTRO; SANTOS; RIBEIRO, 2010). Neste contexto, a correta classificação fiscal de mercadorias e serviços é uma tarefa fundamental, pois determina a carga tributária e a legalidade das operações. Códigos como a Nomenclatura Comum do Mercosul (NCM), que identifica a natureza das mercadorias, e o Código Fiscal de Operações e Prestações (CFOP), que define a natureza da operação, são elementos centrais deste sistema.

Impostos regionais (como por exemplo o ICMS) e impostos que dependem da classificação a partir do NCM (no caso o IPI) é um processo propenso a erros. Tais erros podem acarretar penalidades variando de região para região, além de autuações que afetam a saúde financeira das organizações. O tempo e os recursos despendidos para garantir a conformidade manual representam um custo de transação elevado para as empresas, conforme metodologias de análise sugeridas por órgãos como o CIAT (Centro Interamericano de Administrações Tributárias (CIAT) and Nações Unidas (UN-DESA) 2015).

Para endereçar este problema, propõe-se a aplicação de automação por meio de um sistema de agentes inteligentes. O uso de técnicas de Inteligência Artificial (IA), como o Processamento de Linguagem Natural (PLN), permite que sistemas computacionais processem e interpretem a linguagem humana contida nas descrições de produtos. Modelos de PLN, como o *Bidirectional Encoder Representations from Transformers* (BERT) e suas variantes para o português (Souza; Nogueira; Lotufo 2020), já demonstraram capacidade na compreensão de textos. Recentemente, o uso de Grandes Modelos de Linguagem (LLMs) expandiu as possibilidades, permitindo a extração e a sumarização de regras a partir de vastos volumes de documentos legais.

A solução proposta neste projeto justifica-se pela necessidade de otimizar a gestão fiscal, reduzindo os custos operacionais e os riscos de inconformidade. O sistema combinará agentes baseados em PLN, para a classificação de produtos, com agentes baseados em um Motor de Regras de Negócio (BRE), que automatiza a aplicação da legislação tributária. Essa arquitetura modular oferece uma abordagem funcional e adaptável, capaz de responder às constantes mudanças na legislação, incluindo a recente Reforma Tributária, e de prover uma base para o desenvolvimento de soluções *TaxTech* no Brasil.

Nosso público alvo destina-se aos micro e médio empreendedores assim como analistas contábeis para utilizar essa solução como meio de validação de NF-e



2 Objetivos

2.1 Objetivo Geral

Desenvolver e validar um protótipo de um sistema de agentes inteligentes para automatizar a classificação fiscal de produtos e o cálculo de tributos (ICMS, PIS e COFINS, IPI) em operações comerciais, para auxiliar na validação do profissional de contabilidade em notas fiscais eletrônicas através de um xml (nf-e)

2.2 Objetivos Específicos

1. **Mapear e Modelar Regras Tributárias:** Extrair e estruturar as principais regras fiscais (NCM, IPI, ICMS, PIS/COFINS) a partir de documentos legais, utilizando LLMs para auxiliar na interpretação e sumarização da legislação.
2. **Desenvolver Agente de Classificação (PLN):** Criar e treinar um modelo para aplicar as regras de impostos (ICMS, IPI, PIS e COFINS).
4. **Conectar a APIs Governamentais:** Utilizar do cnpj da NF-e para encontrar o CNAE correspondente e após disso classificar o tipo de segmento daquele cliente,
5. **Validar o Sistema:** Testar o protótipo com um conjunto de dados de NF-e anonimizadas, comparando os resultados com uma base de referência para aferir a precisão do sistema.

3 Metodologia

A metodologia do projeto integrará práticas de ciência de dados e engenharia de *software*, com um ciclo de desenvolvimento iterativo.

1. **Extração e Modelagem do Conhecimento:** A base de conhecimento do sistema será construída a partir da análise de documentos legais, como o Manual de Orientação do Contribuinte da NF-e (Ministério da Fazenda (MF) and CONFAZ

2021) e a legislação pertinente. Grandes Modelos de Linguagem (LLMs) serão empregados em um fluxo de trabalho supervisionado para processar estes documentos, extrair entidades relevantes e sumarizar regras

complexas. O resultado será um conjunto de regras lógicas e parametrizadas, que serão validadas por especialistas antes da implementação.

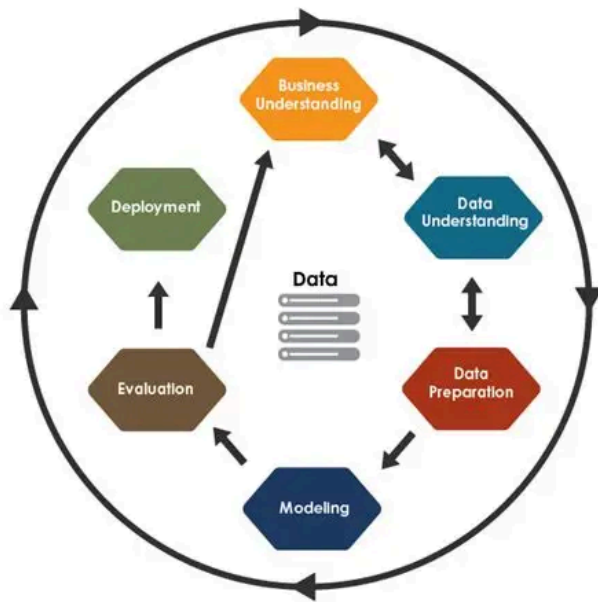
2. Desenvolvimento do Agente de Classificação (PLN):

- **Dados e Pré-processamento:** Será utilizado um *dataset* (conjunto de dados) de NF-e.
- **Modelo de Classificação:** Será implementado um modelo de classificação hierárquica, abordagem que respeita a estrutura em árvore do código NCM trazendo como resposta o IPI correspondente, assim como a partir da verificação da UF-CIDADE origem e destino determinar o ICMS adequado, que demonstrou bons resultados em trabalhos anteriores (A. F. C. d. Silva; S. G. T. d. C. e. Silva 2022). Modelos da arquitetura *Transformer* (BERT) serão utilizados para a representação semântica do texto.

3. **Implementação do Agente de Cálculo (BRE):** Será utilizado um motor de regras como o Drools para codificar a legislação. O motor receberá como entrada os dados do XML NF-e e a classificação sugerida pelo agente de PLN para aplicar as regras e realizar os cálculos.
4. **relatório de inspeção:** Como produto final será entregue um relatório em xlsx para conferência de cada item com o valor de impostos da NF-e e o valor teórico calculado com o intuito de facilitar a interpretação dos valores de impostos analisados (IPI, ICMS, PIS e COFINS)

5 FRAMEWORK

O projeto foi desenvolvido aplicando o framework **CRISP-DM**



C – Business Understanding
(Compreensão do Negócio)

R – Data Understanding (Compreensão dos Dados)

I – Data Preparation (Preparação dos Dados)

S – Modeling (Modelagem)

P – Evaluation (Avaliação)

DM – Deployment (Implantação)

A aplicação do framework CRISP-DM (Cross Industry Standard Process for Data Mining) ao projeto de desenvolvimento do agente inteligente para automação fiscal eletrônica estruturou a construção da solução em etapas metodológicas bem definidas e iterativas.

Na fase de **Compreensão do Negócio (C)**, identificou-se a necessidade de otimizar o processo de análise tributária das Notas Fiscais Eletrônicas (NFes), reduzindo a intervenção humana e mitigando erros decorrentes de cálculos manuais e interpretações inconsistentes de legislações estaduais. O objetivo central foi desenvolver uma ferramenta capaz de automatizar o cálculo de ICMS, PIS, COFINS e IPI, assegurando conformidade fiscal e transparência nos resultados.

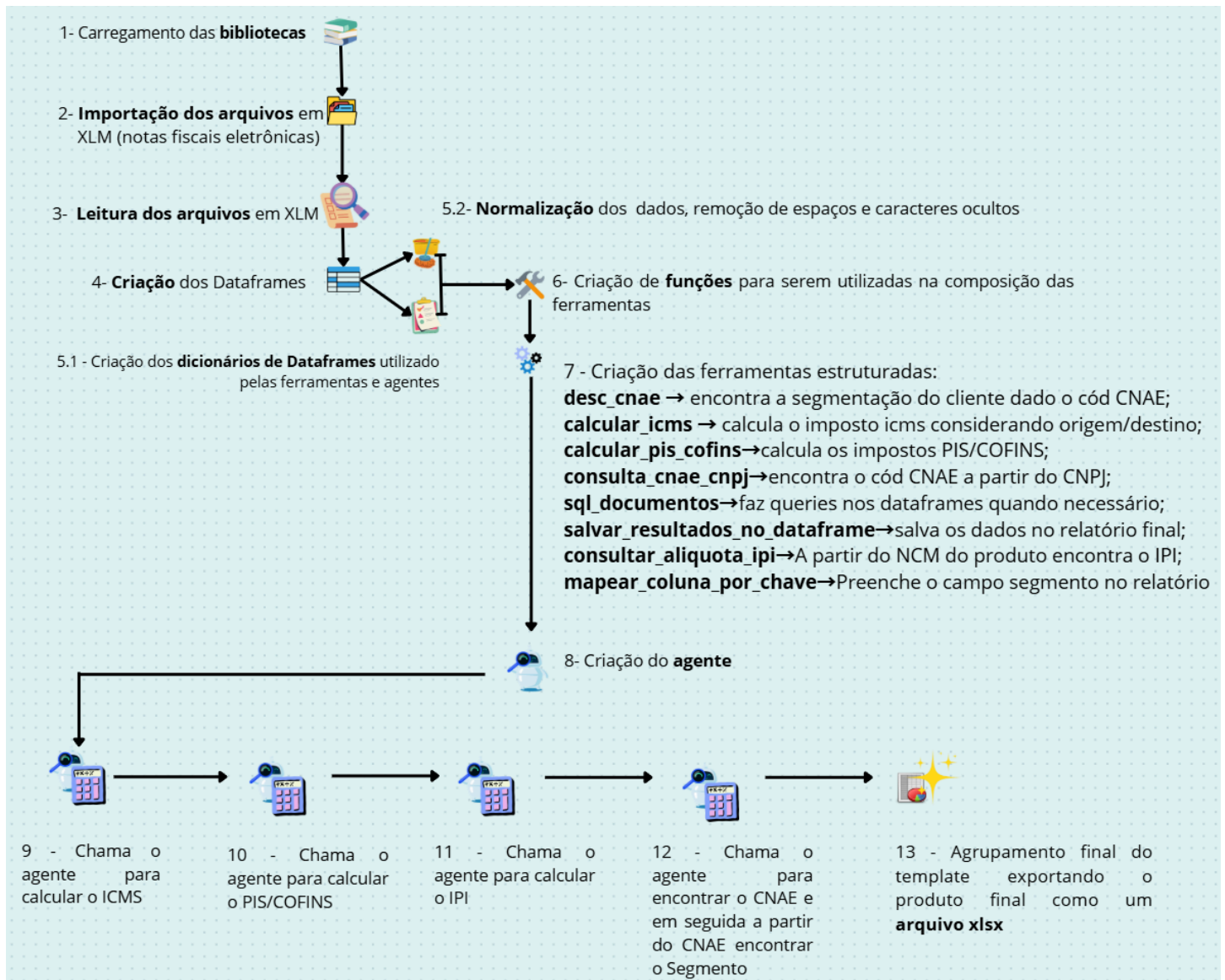
Em seguida, as etapas de **Compreensão dos Dados (R)** e **Preparação dos Dados (I)** concentraram-se na extração, tratamento e padronização das informações provenientes dos arquivos XML das NFes. Essa fase envolveu a normalização de tabelas auxiliares (como NCM, Tabela_IPI, CNAE e alíquotas de ICMS por UF), a correção de formatos heterogêneos e a integração de fontes externas, garantindo consistência e integridade dos dados utilizados no cálculo.

Na fase de **Modelagem (S)**, optou-se pela adoção de um agente autônomo baseado na arquitetura LangChain, estruturado por meio de ferramentas especializadas que representam as regras fiscais de forma modular e interpretável. Cada ferramenta foi projetada para executar funções específicas — como cálculo de tributos, mapeamento de CNAEs e verificação de bases de alíquotas — permitindo interpretação contextual e execução encadeada das operações tributárias.

A etapa de **Avaliação (P)** validou a acurácia dos cálculos automatizados em comparação aos resultados manuais, evidenciando elevado grau de precisão e reprodutibilidade dos valores obtidos. Foram também verificadas a estabilidade do agente, o tempo de execução e a consistência das análises fiscais em diferentes cenários de notas e origens de mercadorias.

Por fim, na fase de **Implantação (DM)**, o agente foi integrado a uma interface interativa desenvolvida em Streamlit, possibilitando o upload de NFes, execução automática das análises e exportação dos relatórios em formato Excel. Essa consolidação transformou o sistema em uma ferramenta de apoio à gestão fiscal, promovendo eficiência operacional, rastreabilidade das decisões e preparando o ambiente para futuras expansões, como integração com ERPs e serviços de verificação tributária em tempo real.

6 FLUXOGRAMA PROCESSO



1. Importação das bibliotecas necessárias;
2. Importação dos arquivos XML através da interface com o usuário;
3. Leitura dos arquivos XML e extração dos dados necessários;
4. Criação dos dataframes:
 - a. **df_cnae** -> contém as descrições e segmentação de todos cnae
 - b. **df_icms** -> core de todas as regras interestaduais icms
 - c. **df_ipi** -> contém todos IPI dado o NCM do produto
 - d. **cnnpj_list** -> utilizado para não ficar repetindo chamada na api somente com cnpjs únicos encontrados no xml para encontrar o cnae correspondente.
 - e. **df_notas** -> template do produto final.
5. Criação do dicionário de dataframes para o agente consumir e normalização dos dados;
6. Criação de funções;
7. Criação das ferramentas utilizadas pelo agente;

- 8. Criação do Agente - modelo gpt4o-mini;
- 9. Chama o agente para realizar o cálculo do **ICMS** considerando as UF de origem e destino do xml;
- 10. Cálculo do PIS/COFINS considerando as alíquotas adequadas, valor bruto da mercadoria e ICMS;
- 11. Pesquisa na tabela qual o IPI% dado o NCM do produto;
- 12. A partir dos CNPJs únicos da lista do XML chama a API do governo para identificar qual o código CNAE e em seguida dado o código CNAE consulta as tabelas internas qual a descrição e segmento do cliente;
- 13. Gera como produto final

Interface Streamlit do agente:

Processamento de Notas Fiscais Eletrônicas - NFe

Faça upload das notas fiscais XML

Drag and drop files here

Limit: 200MB per file • XML

Browse files

xml3.xml

18.1KB

X

8 registros processados de 1 arquivos XML

Pré-visualização das notas carregadas:

	numero_nf	data_emissao	uf	uf_destino	cidade	cnpj	nome_fantasia	ie
0	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
1	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
2	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
3	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
4	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560

Executar agente de análise e cálculo

Iniciando o Fluxo de Cálculos Fiscais...

ICMS calculado e persistido.

PIS e COFINS calculados e persistidos.

Alíquotas IPI consultadas e persistidas.

CNAE e segmento mapeados com sucesso.

Fluxo de análise e ajustes finais concluído com sucesso!

Resultado final (df_notas atualizado):

	numero_nf	data_emissao	uf	uf_destino	cidade	cnpj	nome_fantasia	ie
0	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
1	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
2	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
3	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
4	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
5	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
6	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560
7	7104433	2025-03-03	SC	SP	Joinville	09376495000122	PICHAU JVILLE	2560

Baixar Relatório (relatorio_notas_fiscais.xlsx)

Tabela final:

		nome_fantasia	ie	crt	descricao	ncm	cfop	cest	vProd	icms	pis	cofins	ipi	chave_uf	nome_arquivo	icms_calculado	pis_calculado	cofins_calculado	ipi_calculado	cnae_fiscal	segmento
0	3122	PICHAU JVLE	256042942	3	KR Upgrade AMD 4700S Octa-Core 16 Threads 3.6Ghz (Turbo 4.6Ghz) Memoria RAM 16	85423190	6108	None	974.61	116.95	14.15	65.18	0	SCSP	xml3.xml	116.9532	14.1514	65.1822	0	4751201	COMÉRCIO VARE_JSTA
1	3122	PICHAU JVLE	256042942	3	SSD Mancer Resper RF, 240GB, Sata III 6GB/s, Leitura 530 MB/s, Gravação 500 MB/s, M	84717040	6108	None	203.44	26.79	2.91	13.43	0	SCSP	xml3.xml	24.4128	2.9147	13.4254	0	4751201	COMÉRCIO VARE_JSTA
2	3122	PICHAU JVLE	256042942	3	Placa de Video Mancer Radeon RX 550, 4GB, GDDR5, 128-bit, MCR-RX550-4GBV3	85423190	6108	None	471.39	57.7	6.83	31.44	0	SCSP	xml3.xml	56.5668	6.8259	31.4404	0	4751201	COMÉRCIO VARE_JSTA
3	3122	PICHAU JVLE	256042942	3	Fonte TGT Tomahawk T3, 600W, Preto, TGT-T2WK600-BK01	85044021	6108	None	124.25	15.47	1.79	8.27	0	SCSP	xml3.xml	14.91	1.7949	8.2673	0	4751201	COMÉRCIO VARE_JSTA
4	3122	PICHAU JVLE	256042942	3	Gabinete Gamer Mancer Hexer, Rainbow, Mid-Tower, Lateral de Vidro, Com 3 Fans, Pn	84733019	6108	None	170.09	21.74	2.45	11.27	0	SCSP	xml3.xml	20.4108	2.4478	11.2746	0	4751201	COMÉRCIO VARE_JSTA
5	3122	PICHAU JVLE	256042942	3	Cabo HDMI Pichau, 1.8m, Preto, PG-HDTV-BL01	85444200	6108	None	5.29	0.67	0.08	0.35	0	SCSP	xml3.xml	0.6348	0.0762	0.3511	0	4751201	COMÉRCIO VARE_JSTA
6	3122	PICHAU JVLE	256042942	3	Cabo de Força Pichau, 1.5m, 3 Pinos, Preto, PG-PWC-BL01	85444200	6108	None	5.29	0.67	0.08	0.35	0	SCSP	xml3.xml	0.6348	0.0762	0.3511	0	4751201	COMÉRCIO VARE_JSTA
7	3122	PICHAU JVLE	256042942	3	KIT Ventilinha Aceageek Crystal, 3x120mm, ARGB, Preto, AG-CRYSTAL3X1-BK	84145990	6108	None	91.99	11.04	1.34	6.15	0	SCSP	xml3.xml	11.0388	1.3357	6.1522	0	4751201	COMÉRCIO VARE_JSTA

Conclusão

O desenvolvimento e a implementação do agente autônomo para automação do cálculo tributário demonstraram a viabilidade técnica e operacional da utilização de tecnologias de inteligência artificial aplicada à análise fiscal. A adoção da arquitetura baseada em agentes, aliada ao framework LangChain e à integração com bases estruturadas (NCM, IPI, CNAE e alíquotas de ICMS), possibilitou a criação de um sistema modular, transparente e escalável, capaz de executar cálculos complexos com elevado grau de precisão e rastreabilidade.

Os resultados evidenciaram consistência entre os valores automatizados e os cálculos manuais, confirmando a conformidade das operações fiscais e a robustez da abordagem proposta. Além disso, a utilização do framework CRISP-DM como metodologia de desenvolvimento garantiu uma estrutura sistemática para o processo, desde a compreensão do domínio tributário até a implantação da solução em ambiente de uso real.

Do ponto de vista técnico, o agente demonstrou eficiência na manipulação e interpretação de dados em XML, reduzindo significativamente o tempo de processamento e a possibilidade de erro humano. A integração com a interface Streamlit reforçou o caráter prático da ferramenta, viabilizando a interação direta com o usuário e a exportação automatizada de relatórios fiscais padronizados. Em síntese, o projeto consolidou-se como uma prova de conceito sólida para a aplicação de agentes inteligentes em processos fiscais automatizados, apresentando potencial de expansão para outros domínios da contabilidade e compliance tributário. Futuras evoluções podem incluir o uso de modelos de linguagem mais avançados, integração com APIs governamentais em tempo real e o aprimoramento de mecanismos de auditoria e explicabilidade dos cálculos realizados.

projeto disponibilizado em:

[agentes_do_amanha/trabalho_conclusao_2](https://github.com/PV-Alcantara/agentes_do_amanha/tree/main) [at](#) [main](#) .
[PV-Alcantara/agentes_do_amanha](#)

link para utilização do app:

<https://pv-alcantara-agentes-d-trabalho-conclusao-2streamlit-app-3zi4uw.streamlit.app/>