

IAGO ARAGÃO

Modelagem e Dados

📍 Teresina, PI ✉ iagoaragao88@gmail.com ☎ +55 (86) 99485-6119
🌐 [linkedin.com/in/iago-aragao](https://www.linkedin.com/in/iago-aragao) 🐙 github.com/iago-aragao 🌐 Portfólio: iago-aragao.github.io

PERFIL

Profissional de Dados com experiência na área desde 2022, incluindo atuação empresarial como Analista de Dados e atuação atual como Pesquisador em Machine Learning. Atua da análise, preparação e qualidade dos dados à construção, avaliação e publicação de modelos estatísticos, soluções de Machine Learning e produtos analíticos em Python e SQL, com atenção à interpretação de resultados, confiabilidade técnica e contexto de negócio.

HABILIDADES TÉCNICAS

Análise & Modelagem: Python, SQL, Pandas, análise exploratória, análise de dados, modelagem estatística, modelagem analítica

Machine Learning: Scikit-learn, XGBoost, classificação, validação temporal, avaliação de modelos, métricas de classificação, dados desbalanceados, interpretação de modelos

Engenharia & Qualidade de Dados: Airflow, DuckDB, ETL, ingestão e transformação de dados, qualidade de dados, pipelines de dados, orquestração, data marts

Cloud & BI: AWS, Power BI, análise e visualização de dados

Desenvolvimento & Ferramentas: FastAPI, Streamlit, Git, APIs REST, Docker, testes automatizados, CI, observabilidade, logs

Idiomas: Inglês avançado

EXPERIÊNCIA

Orientador de Informática

Sociedade Espírita João Nunes Maia

Fev 2026 – Ago 2026

Teresina, PI, Brasil

Atuação profissional em projeto social apoiado pela [Fundação Francisco e Clara de Assis](#)

- Ministro aulas de informática básica para crianças de baixa renda em iniciativa de inclusão digital, com foco no desenvolvimento de habilidades digitais fundamentais.
- Apoio atividades de inclusão e empreendedorismo digital, orientando crianças e adolescentes no desenvolvimento de competências tecnológicas essenciais.

Pesquisador em Machine Learning — Universidade Federal do Piauí

Pesquisa aplicada — Programa PIBITI — Remoto

2025 – Atual

Teresina, PI, Brasil

- Desenvolvo experimentos de Machine Learning aplicado, cobrindo preparação de dados, treinamento, avaliação estatística e validação de resultados.
- Pesquiso modelos para visão computacional e NLP, incluindo inspeção industrial com *one-class learning* e arquiteturas híbridas Transformer-BiLSTM.
- Construo protótipos que combinam modelos preditivos e LLMs para apoio à decisão, análise de risco e automação com revisão humana.

Analista de Dados — Ivan Casa e Construção

Presencial

Fev 2022 – Abr 2023

Teresina, PI, Brasil

- Organizei e validei dados fiscais e de produtos no ERP, garantindo informações consistentes para rotinas tributárias e operacionais.
- Identifiquei e corriji erro crítico de parametrização fiscal por meio de análise exploratória de dados, reduzindo falhas de cálculo e riscos de conformidade.
- Criei dashboard no Power BI para acompanhar dados fiscais e de produtos e automatizei validações tributárias, reduzindo em 80% o tempo de auditoria manual.

PROJETOS EM DESTAQUE

FraudRisk Engine | Machine Learning & Backend

2026

- Desenvolvi sistema de score de fraude com XGBoost e FastAPI, avaliado em holdout temporal de 42.722 transações históricas do dataset público ULB/OpenML. Detectou 44 de 52 fraudes (84,6% de recall) e 78,0% do valor fraudulento, encaminhando 0,817% das transações para revisão. [GitHub](#)

Airflow Data Quality Pipeline | *Engenharia de Dados & FastAPI*

2026

- Construí pipeline de dados de varejo com Airflow, DuckDB e FastAPI, cobrindo ingestão, validação, transformação e publicação. Processou 541.909 registros de origem, aprovou 25/25 verificações de qualidade e gerou 4 marts analíticos; possui publicação atômica, bloqueio em falhas críticas, 35 testes automatizados e 94% de cobertura. O DAG registra logs de execução e aplica retries nas tarefas críticas; as saídas são publicadas pela API somente após as validações. [GitHub](#)

QueryGuard RAG | *Text-to-SQL & Segurança*

2026

- Desenvolvi backend Text-to-SQL com catálogo semântico e recuperação lexical de esquema, incluindo políticas por função, bloqueio de dados sensíveis, validação via `sqlglot`, execução somente leitura e auditoria. Alcançou 50/50 casos na avaliação determinística e 5/5 no fallback com LLM, com 72 testes automatizados. [GitHub](#)

AI Ops Approval Workflow | *IA Generativa & Automação*

2026

- Desenvolvi backend para triagem de solicitações operacionais com FastAPI e SQLite, incluindo aprovação humana, regras determinísticas de segurança e trilha de auditoria. Integra opcionalmente a OpenAI Responses API com Structured Outputs e fallback seguro, além de disponibilizar workflow n8n importável. [GitHub](#)

Property Value Insights | *Machine Learning & MLOps*

2026

- Desenvolvi sistema de estimativa de preços residenciais com HistGradientBoostingRegressor, validação temporal em cinco janelas e API FastAPI pública. Treinado com 21.595 vendas, alcançou MAE médio de US\$ 63.880,80; inclui demonstração web, imagem Docker, verificação SHA-256 do artefato, logs estruturados e métricas Prometheus. [GitHub](#) | [Demo](#)

PESQUISA E ARTIGOS

I Speak Kanoê | *SBBB 2026 DSW — Artigo aceito*

2026

- *I Speak Kanoê: Engineering a Dataset for an Endangered Isolated Language*. Coautor de pesquisa sobre dataset público para a língua indígena Kanoê, com foco em NLP para línguas de baixo recurso, preservação cultural e tecnologias inclusivas, com publicação prevista nos anais do SBBB 2026. [GitHub](#) | [HuggingFace](#)

Modelos Lexicais Calibrados | *SBBB 2026 Short Papers — Artigo aceito*

2026

- *Modelos Lexicais Calibrados para Detecção de Ofensividade em Português Brasileiro: um Estudo no ToLD-Br*. Pesquisa sobre detecção de ofensividade em português brasileiro, com publicação prevista nos anais do SBBB 2026.

EDUCAÇÃO

Universidade Federal do Piauí
Curso Superior de Tecnologia em Gestão de Dados – IRA: 9.31/10

Conclusão prevista: 1º semestre de 2027
Teresina, PI, Brasil

CERTIFICAÇÕES

AWS Technical Essentials <i>AWS Training and Certification</i>	2026
AWS Cloud Practitioner Essentials <i>AWS Training and Certification</i>	2026
AWS Educate Introduction to Cloud 101 <i>AWS Training and Certification</i>	2026
AWS Educate Getting Started with Storage <i>AWS Training and Certification</i>	2026
Introdução à Análise de Dados – Microsoft Power BI <i>Fundação Bradesco</i>	2026
CS50's Introduction to Computer Science <i>HarvardX</i>	2026
Deep Learning com Python <i>UFPI</i>	2024
Desenvolvimento de Games 3D <i>UFPI</i>	2024