

Davi Specia

📍 Toledo, PR, Brasil ✉️ davi.specia@gmail.com ☎️ (45) 99128-5227 🌐 davispecia.com.br 🌐 davispecia
🔗 davidogral

Resumo

Trabalho com dados em operação real e com recuperação de informação. Sou sócio-fundador da Alugaix, onde lidero a equipe técnica e construí a camada analítica de uma gestora de 60 imóveis. Fora dela, mantenho um sistema de busca semântica em produção, um pipeline de dados de crédito no Databricks e uma rede neural implementada do zero em Rust, com artigo aprovado no Latinoware. Procuo agora uma vaga de analista de dados em uma equipe onde eu seja liderado e ganhe senioridade com gente mais experiente do que eu. Python, SQL e Rust.

Experiência

Alugaix, Sócio-fundador e analista de dados

Toledo, PR

Mar 2024 – presente

- Responsável pela área de dados de uma operação com 60 imóveis sob gestão e faturamento anual acima de R\$ 1,5 milhão em Toledo e região. Integro o PMS (Hostify), a precificação (PriceLabs) e os canais de venda (Airbnb, Booking).
- Construí a camada analítica da operação, com ingestão via API, padronização dos dados e indicadores de ocupação, ADR e RevPAR, usados hoje no fechamento mensal com os proprietários.
- Desenvolvi o Compacta, ferramenta que cria anúncios a partir das fotos e dos documentos do imóvel. Um modelo de visão classifica as fotos por cômodo, LLMs extraem os dados dos documentos e escrevem os textos, e o anúncio é publicado direto na Hostify pela API. Cada anúncio leva pelo menos 5 horas a menos para ficar pronto. Em produção com FastAPI e Docker. (sistema interno)
- Como sócio-fundador, lidero uma equipe multidisciplinar e transformo as necessidades do negócio em arquitetura técnica e roadmap de produto.

Projetos

Cinerd, busca de filmes por descrição (Ver projeto)

- Sistema de busca em produção que encontra um filme a partir de uma descrição vaga do usuário. O catálogo tem 22.029 títulos da API do TMDB, completados com a base pública do IMDb e o enredo da Wikipédia.
- A busca combina 8 sinais, entre eles BM25 e embeddings multilíngues (multilingual-e5-large) sobre sinopse e enredo, nome de personagem com tolerância a erro de digitação e popularidade. Os scores são normalizados por z-score com limite para valores extremos.
- No conjunto de teste, separado do usado na calibração, o nDCG@10 chegou a 0,829, contra 0,360 do BM25 sozinho. Na ablação, a maior melhora veio do limite aplicado ao z-score.
- Um LLM (Groq) reescreve a consulta antes da busca e depois confere os 20 primeiros resultados lendo a sinopse. Nas consultas mais difíceis, o nDCG@10 subiu de 0,473 para 0,654.
- Testei um cross-encoder e tirei da produção, porque o ganho ficou dentro do ruído e a latência aumentou cerca de 40 vezes. A quantização INT8 com ONNX Runtime reduziu a memória de 2,3 para 1,7 GB quase sem perda de qualidade.
- Deploy próprio na Oracle Cloud, com dados e índices versionados em DVC, avaliação automatizada, CI com ruff, mypy e pytest e métricas por etapa no formato Prometheus.
- Recomendação híbrida com SVD próprio (64 fatores) treinado em 10 milhões de avaliações do MovieLens, com RMSE de 0,83, combinado por RRF com similaridade de conteúdo.

Pipeline de análise de crédito (Ver projeto)

- Pipeline em arquitetura medalhão rodando no Databricks como job serverless de 6 tarefas, com infraestrutura como código (Databricks Asset Bundles), tabelas Delta no Unity Catalog e carga incremental com MERGE INTO.
- Quality gate com 20 regras que bloqueia a publicação e as tarefas seguintes (analytics e modelo em pyspark.ml), com cada execução registrada numa tabela de auditoria.
- A mesma lógica roda localmente com PostgreSQL e Airflow, com mais de 40 testes automatizados e CI no GitHub Actions.

Rede neural guiada por física para Itaipu (Ver projeto)

- Implementei a rede do zero em Rust, sem bibliotecas de machine learning, com retropropagação e otimizador Adam escritos à mão e conferidos por gradient check. Dados do ONS de 2015 a 2024, com 2023 e 2024 reservados para teste.
- Na ablação, o modelo sem o termo físico teve o menor erro nos dados, mas um resíduo físico cerca de 33 vezes maior (0,559 contra 0,017). Ou seja, errar menos nos dados não garantiu um modelo coerente com a física.
- Usei a persistência (geração do dia anterior) como baseline. A vazão sozinha melhora pouco sobre ela (RMSE de 586 contra 590 MWmed), mas o cenário operacional com MLP residual chegou a 508 MWmed (R^2 de 0,91) e superou a persistência em todos os anos da validação móvel de 2020 a 2024.
- Artigo aprovado no Latin.Science 2026 (Latinoware). Os resultados se reproduzem a partir de um clone limpo, com CI rodando clippy, testes e checagem SHA-256 dos dados.

Publicações

Estimativa de peso de suínos a partir de imagens RGB utilizando redes neurais convolucionais e o dataset PigRGB-Weight

Ago 2026

Vinicius Tessele, Jhonatan da Silva Margraf, Vinicius Hillebrand Carraro, Marco Antonio Wandraski, Luiz Henrique Zavatini Feltrin, Davi Specia
RECIMA21, v. 7, n. 8, DOI [10.47820/recima21.v7i8.8496](https://doi.org/10.47820/recima21.v7i8.8496)
recima21.com.br/recima21/pt_BR/article/view/8496

Modelagem da Geração Hidrelétrica de Itaipu via Redes Neurais Guiadas por Física em Rust: Validação Temporal, Persistência e Previsão Operacional

2026

Davi Specia, Luiz Henrique Zavatini Feltrin, Tainá Dreissig, Miguel Mantoan Castellani, Vinicius Carraro
Latin.Science 2026, Latinoware, artigo aprovado · [Ler artigo no Drive](#)

Formação

Faculdade Donaduzzi (Biopark), Bacharelado em Inteligência Artificial

Toledo, PR

- Os primeiros 2,5 anos seguem a mesma grade do curso de Ciência de Dados, o que permite obter também esse diploma.

Fev 2025 – Dez 2029

IFPR, Campus Toledo, Técnico em Informática para Web integrado ao Ensino Médio

Toledo, PR

Fev 2022 – Dez 2025

Competências

Linguagens: Python, SQL, Rust, Dart

Dados e análise: pandas, NumPy, scikit-learn, XGBoost, estatística descritiva e inferencial, modelagem de bases, ETL

NLP e recuperação: sentence-transformers, embeddings multilíngues, BM25, busca híbrida, fuzzy matching, LLMs

Deep learning: PyTorch, TensorFlow/Keras, visão computacional

MLOps: Databricks, PySpark, Airflow, DVC, MLflow, Docker, GitHub Actions, Prometheus, Grafana, ONNX Runtime

Serviço e visualização: FastAPI, Flask, Power BI, Streamlit, Dash

Idiomas

Português: nativo

Inglês: básico, leitura de documentação técnica