PEDRO JULIUS LEMOS DOS SANTOS

Engenheiro de IA | AI Engineer

Machine Learning, Deep Learning e Engenharia e Análise de Dados

Fortaleza, Ceará, Brasil | +55 (85) 98553-6293 | pdrjuliusls@gmail.com

pedrojulius.tech | linkedin.com/in/pedro-julius | github.com/pjuliusl

Perfil profissional

Profissional de IA aplicada e dados, com atuação como Analista de Dados na Polícia Civil do Ceará. Desenvolvo pipelines em produção, modelos analíticos e integrações entre aplicações de IA e ferramentas de dados. Atuo com Python, SQL, Airflow e dbt, além de MCP, fine-tuning com LoRA e avaliação de modelos com MLflow. Minha experiência inclui qualidade e rastreabilidade de dados, testes automatizados e revisão humana de respostas de IA.

Competências técnicas

IA e MLOps: PyTorch, Transformers, PEFT/LoRA, MLflow, LiteLLM, MCP, fine-tuning, avaliação e calibração de modelos.

Engenharia de dados: Python, SQL, Apache Airflow, dbt, PostgreSQL, ClickHouse, pipelines ETL/ELT, validação e qualidade de dados.

Analytics e BI: Apache Superset, Power BI, modelagem de dados, indicadores operacionais e análise de resultados.

Desenvolvimento e entrega: FastAPI, Pydantic, SQLAlchemy, Alembic, pytest, Git, Docker, Jenkins e Jobs de treinamento em Kubernetes/K3s.

Integrações seguras: APIs REST, OAuth2/OIDC, JWT, controle de acesso, proteção CSRF e sessões com Redis/Sentinel.

Experiência profissional

Coordenadoria de Tecnologia e Informação da Polícia Civil do Ceará

Analista de Dados | atuação terceirizada | jul. 2025 - atual

- Projeto e mantenho DAGs no Airflow e transformações com dbt, SQL, PostgreSQL e ClickHouse, investigando falhas e evoluindo pipelines em produção.
- Corrigi uma lacuna de captura incremental que deixava alterações em registros antigos fora do destino e confirmei o resultado no banco final.
- Implementei extração de indicadores de PDFs semiestruturados em Python/Airflow, com validação, rastreabilidade e comparação mensal de duas fontes em painéis Superset.
- Desenvolvo integrações MCP para o Superset e soluções de treinamento, avaliação e revisão humana de modelos de IA.

iREDE — Instituto de Gestão, Redes Tecnológicas e Energias

Assistente de Projetos | out. 2024 - jun. 2025

- Acompanhei entregas e produzi documentação técnica de projetos.

Monitor de Desenvolvimento Android | out. 2023 - dez. 2024

- Preparei aulas práticas e trilhas de Kotlin e acompanhei alunos iniciantes no desenvolvimento de aplicações Android.

Escola Municipal Maria Bezerra Quevedo

Estagiário de Suporte de TI | ago. 2022 - fev. 2023

- Prestei suporte na sala de informática, com instalação de software, manutenção de equipamentos, operação de Chromebooks e elaboração de relatórios.

Projetos em IA aplicada e Machine Learning

Integração MCP com Apache Superset | PCCE

- Entreguei a fundação de um MCP Host/Client em Python para integrar um assistente analítico ao Superset, com autenticação, sessões e validação do catálogo de ferramentas.
- Contribuí com o runtime MCP do Superset, incluindo validação JWT/JWKS, escopos e restrição das ferramentas autorizadas, com testes de autenticação e dos contratos de integração.
- Implementei a inicialização do chat no dashboard com proteção CSRF e validação de acesso. Separei sessões e nonces por usuário, sessão e aba, usando Redis com expiração e operações atômicas.
- Adicionei suporte a Redis Sentinel e leitura de segredos em containers. Preparei builds e smoke tests no Jenkins e publicação de imagens no Harbor.

Fine-tuning e avaliação de modelos de decisão | PCCE

- Implementei um pipeline de fine-tuning LoRA para classificação estruturada em quatro categorias, usando PyTorch, Transformers e PEFT sobre um modelo de linguagem pré-treinado.
- Preparei conjuntos de treino, validação, calibração e teste final, com agrupamento por registro para evitar contaminação entre partições e preservar a avaliação independente.
- Preparei Jobs de treinamento em GPU no Kubernetes/K3s, com imagens de dependências fixadas, checkpoints, retomada de execução e controle de interrupção do treinamento.
- Integrei MLflow para registrar parâmetros, métricas e referências de versão. Mantive o conteúdo dos registros fora dos logs e do rastreamento de experimentos.
- Implementei avaliação por macro-F1, acurácia balanceada, precisão e recall por classe, além de critérios de seleção de checkpoints que limitam erros graves e apoiam a calibração.
- Desenvolvi exportação dos pesos com verificações de paridade entre treinamento e inferência, identificação da release e preparação de candidatos para avaliação antes da produção.

Plataforma de revisão humana de respostas de IA | PCCE

- Desenvolvi aplicações de revisão supervisionada com FastAPI, SQLAlchemy, PostgreSQL e Alembic, com fila de trabalho, rascunhos, justificativas e rastreabilidade das decisões.
- Implementei revisão cega e comparação entre respostas de modelos e analistas, separando perfis de acesso e protegendo os fluxos de revisão com autenticação e CSRF.
- Construí painéis de concordância, erros graves, cobertura e confiança, além de indicadores operacionais de volume, falhas e latência para acompanhar a avaliação dos modelos.
- Adicionei importação de revisões, exportação de resultados e testes automatizados. Estruturei o uso de dados simulados para validar os fluxos antes da integração operacional.

Chat Infinito, fork do UniiChat | Projeto pessoal

- Adaptei o UniiChat, chat de código aberto de Victor Taelin que guarda a conversa inteira numa árvore de resumos, para rodar sobre as assinaturas Claude e ChatGPT pelos clientes oficiais de linha de comando, sem chaves de API.
- Implementei resumos em lote pelo Codex, troca de modelo quando um plano atinge o limite, busca exata no histórico e uma regra que preserva links e números nos resumos, com testes de ponta a ponta.
- Adicionei a exibição do uso real das assinaturas, o isolamento entre integrações e um serviço systemd para uso diário. Stack: TypeScript, Bun, MCP.

Projetos em dados e automação

Dashboard analítico em Power BI | Portfólio

- Desenvolvi um dashboard Power BI conectado ao PostgreSQL usando a base demonstrativa Northwind, com análises de vendas, receita por produto, estoque e ticket médio.
- Organizei os indicadores para exploração dos resultados e comparação de produtos, aplicando visualização de dados a um cenário de negócio demonstrativo.

Formação acadêmica

Bacharelado em Engenharia de Telecomunicações

Instituto Federal do Ceará — IFCE | jul. 2023 - conclusão prevista em jul. 2028

Graduação em andamento na área de sistemas de telecomunicações.

Técnico Integrado em Informática

Instituto Federal do Ceará — IFCE | jul. 2018 - jul. 2022 | Concluído

Formação técnica em informática integrada ao ensino médio.

Formação complementar

Introdução ao Desenvolvimento Android

Fundamentos do desenvolvimento de aplicações móveis Android.

Iniciando a carreira de desenvolvedor Front-end

Fundamentos para o desenvolvimento de interfaces web.

Idiomas

Português: Nativo.

Inglês: EF SET 82/100 — C2 Proficient.