

Skills

- Python (Pandas, NumPy, SciPy, Torch, TensorFlow, Sklearn, Shap, Transformers, OpenCV, LightGBM, XGBoost, LightGBM, CrewAI, Langchain, LangGraph, Plotly, Seaborn, MatPlotLib, NLTK, FastAPI)
- PySpark
- ETL
- NLP
- RAG
- SQL (SQL Server, MySQL, PostgreSQL)
- noSQL(MongoDB, Redis)
- VectorDB (PineCone, cosmosDB, Qdrant , Chroma)
- Docker, Docker compose, Portainer
- LLMs(BERT, T5,GPT, Llama, gemini)
- CNNs(YOLO, VGG, Resnet)
- Microsoft Azure (DataBricks, Azure Data Lake, Azure Data Warehouse, Azure ML, Azure Form recognizer, Blobstorage)
- GIT, Github, GitLab, Azure devops
- OpenAI API, Azure OpenAi, Hugging Face
- Airflow, Minio, Nessie, Trino, Mlflow, Superset

Work Experience

SR DATA SCIENTIST | SR CONSULTANT – KPMG Brasil

October 2022 -

- Developed computer vision systems for multiple domains, including biological inventory automation for large farms, vehicle (cars and buses) inventory tracking, and vessel risk-zone identification, building and training custom datasets and deep learning models to improve operational accuracy and reduce manual inspection effort by up to ~90%.
- Built NLP and LLM-based solutions using GPT for information retrieval over large document sets, automatic summarization, document-based Q&A bots, and financial consolidation data structuring, improving access to critical business information and reducing manual analysis effort.
- Developed end-to-end RAG, OCR, and LLM pipelines for large-scale unstructured data processing, enabling semantic search, document understanding, and structured extraction from heterogeneous sources in enterprise environments.
- Built multi-agent systems using LangGraph and LangChain, integrating Retrieval-Augmented Generation (RAG), web scraping, mathematical tool usage, and reasoning workflows to automate complex decision-making and information processing tasks.
- Implemented a financial data mining system that reduced document processing costs by ~95%, leveraging scalable Python-based pipelines, automated extraction workflows, and structured data transformation processes.
- Performed data quality assessment and exploratory data analysis (EDA) across multiple clients, designing predictive models and analytical solutions that optimized internal processes and reduced operational costs.
- Led early-stage project planning and requirements gathering in direct collaboration with clients, translating business needs into scalable technical solutions and ensuring alignment across stakeholders.
- Supported and mentored junior engineers through task breakdowns, code reviews, and technical guidance, while also delivering internal training sessions on Machine Learning, regression/classification pipelines, and generative AI applications.

RESEARCHER – UNIFOR Universidade de Fortaleza

August 2017 - September 2022

- Developed and led research projects at UNIFOR focused on computer vision, machine learning, and data science, with

emphasis on unstructured data mining and automated information extraction from complex document sources.

- Built end-to-end solutions for structuring and extracting information from unstructured data, reducing manual effort and enabling scalable document intelligence pipelines for real-world applications.
- Delivered applied research systems for public and private sector organizations, including MPCE, UNIMED, and KPMG, translating academic research into production-oriented machine learning solutions for document processing and analytics.

MACHINE LEARNING DEVELOPER – KnowCode

January 2020 - September 2022

- Developed a computer vision and machine learning system for understanding and extracting information from design artifacts (including images and Figma files), enabling automated identification of UI components and their properties.
- Built an end-to-end service that structures extracted design information into a machine-readable format, enabling automatic generation of frontend applications from UI designs and significantly reducing manual implementation effort.

Education

MASTER COMPUTER SCIENCE – Universidade of Fortaleza - UNIFOR

July 2022- incomplete

BACHELOR COMPUTER ENGINEERING – Universidade of Fortaleza - UNIFOR

June 2022

MECHATRONICS TECHNICIAN – SENAI

June 2017

Language

- **ENGLISH: C2**
- **PORTUGUESE: NATIVE**

Projects

USE OF COMPUTER VISION FOR CLAIMS ADJUSTMENT AUTOMATION – Personal Project

june 2022

- Developed an end-to-end computer vision system to automate medical claims auditing for a large health insurance provider, extracting structured information from scanned medical forms and reducing manual review workload.
- Built and curated a labeled dataset of ~2,000 medical form images, including stamps, signatures, and service descriptions, applying data augmentation and expert validation to improve model robustness.
- Implemented a multi-model pipeline combining YOLOv5 for object detection and OCR (Amazon Textract, Tesseract) with string similarity (Levenshtein distance) for robust information extraction and validation.
- Designed a full auditing automation workflow that compares extracted medical services with internal provider systems, automatically flagging non-compliant or non-prescribed procedures with ~94.56% extraction accuracy across 2,600+ detected fields.

DATA MINING IN FINANCIAL DOCUMENTS – KPMG

February 2023

- Developed an end-to-end financial document data mining system for processing 10+ years of invoices in a road construction audit project, enabling structured extraction and analysis of over 92,000 documents.
- Built an automated invoice processing pipeline using Python, Azure Form Recognizer OCR, regex-based access key extraction, and LLM/NLP workflows (LangChain) to extract key financial fields such as CNPJ, issuer name, invoice value, date, and metadata.
- Designed a classification architecture to separate product and service invoices, integrating OCR modules, document readers, and rule-based + API-driven processing flows (including SERPRO integration) to generate structured reports in CSV/JSON formats.
- Delivered a scalable reporting and analytics pipeline that processed 88,000+ valid invoices, achieving ~86% data extraction accuracy and reducing processing time from 48 weeks to 3 weeks, cutting operational costs by over 90%.

- Developed a computer vision system for road feature extraction from satellite imagery, enabling segmentation of streets and sidewalks to support urban mapping and infrastructure analysis.
- Built a full ML pipeline using 18,000+ satellite images and OpenStreetMap data, including geospatial overlay processing, polygon extraction, and dataset construction for supervised learning tasks.
- Trained classification and segmentation models to validate road polygon alignment and predict road structures, improving dataset quality through automated filtering and manual annotation workflows.
- Designed post-processing pipelines using polygon simplification, curve fitting, and graph-based road reconstruction to connect segments and extract navigable road networks with intersections and connectivity logic.