MARCELO PRATES, PHD

Data Scientist & Machine Learning Engineer

@ marceloorp@gmail.com   in marceloprates   ⌗ marceloprates

EXPERIENCE

Lead Data Scientist

Dataside

📅 Oct 2023 – Now   📍 São José dos Campos, Brazil (Remote)

- Designed and implemented a large-scale enterprise LLM provider for a global company.
- Developed and deployed machine learning solutions for various applications, including photogrammetry for 3D reconstruction, computer vision for object detection and image segmentation, and OCR for automated text extraction from images. Implemented advanced algorithms for sales demand forecasting, classification, and regression tasks, leveraging techniques such as time series analysis, ensemble methods, and deep learning models.
- Implemented advanced techniques such as Retrieval-Augmented Generation (RAG), Conversational AI Generation (CAG), Text-to-SQL, and multi-agent systems to enhance the capabilities and performance of machine learning models.
- Mentored 5 junior data scientists.
- Tech stack: Pytorch, Pytorch Lightning, Pandas, Scikit-learn, Numpy, Jupyter Notebooks, Azure OpenAI, Azure Databricks, HuggingFace Transformers, LangChain, LangGraph, BertTopic, CrewAI

AI/LLM Consultant

Vortigo

📅 Nov 2023 – Jun 2024   📍 Porto Alegre, Brazil (Remote)

- Designed and implemented advanced assistant ChatBots leveraging OpenAI's GPT-3.5 and GPT-4, enhancing customer service automation with proprietary source code and spreadsheet knowledge bases.

Sabbatical Period

for Generative Art and Open Source Development

📅 Apr 2023 – Oct 2023   📍 Porto Alegre, Brazil

- Dedicated time to advance my generative art projects, resulting in the creation of several innovative pieces.
- Enhanced and maintained my open-source Python package, prettymaps, which has been well-received by the community for its utility in artistic and data visualization projects.
- Utilized this period to refine my skills in Python and creative coding, contributing to my overall growth as a data scientist and developer.

Computer Vision Consultant

ConstructIN

📅 Mar 2022 – Apr 2023   📍 Porto Alegre, Brazil (Remote)

- Designed and implemented cutting-edge computer vision solutions for automated construction site monitoring using 360-degree photography.
- Led a team of 3 data scientists in developing 4 innovative CV applications, including a dashboard for clients to monitor construction progress, powered by a state-of-the-art semantic segmentation module.

Generative Art Instructor

Responsive Cities Workshop

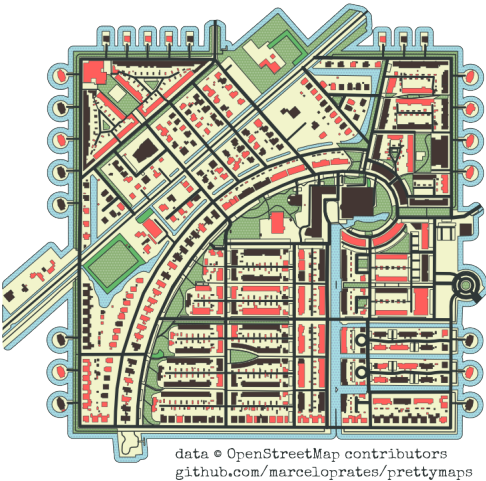📅 Nov 2022   📍 Porto Alegre, Brazil

- Created and taught a course on generative art, covering its history and key concepts.
- Gave practical training with popular tools and libraries, helping participants make their own generative art projects.

MOST PROUD OF

Prettymaps
Python library for drawing pretty maps from OpenStreetMap data (1st place on Hacker News, >11.000 stars, ranked among the 2000 most starred repositories on GitHub)

data © OpenStreetMap contributors
github.com/marceloprates/prettymaps

VO2 Estimation in Samsung Galaxy Watch
Led the development of the VO2 estimation feature in Samsung Galaxy Watch, now shipped globally starting from Galaxy Watch 4

Gender Bias in Machine Translation
Our 2018 research measured the prevalence of male defaults in Machine Translation tools (>500 citations)

Erdős Number = 4
My Erdős number is 4, through Moshe Vardi

LANGUAGES & LIBS

Python	●●●●●
C, C++	●●●○○
C#	●●●○○
Haskell	●●●●○
Julia	●●●●○
Java	●●●○○
Pytorch	●●●●●
Tensorflow	●●●○○
Langchain, LangGraph	●●●●●
CrewAI	●●●●○
HuggingFace Transformers	●●●●●
Pandas, Sklearn, Numpy	●●●●●
SQL	●●●●●
Azure Databricks, MLflow	●●●●●
FastAPI	●●●●○
OpenCV, Open3D, Skimage	●●●●●
Docker, Kubernetes	●●●○○

PROFICIENCY

Portuguese (native speaker)	●●●●●
English	●●●●○
Spanish	●●●○○

Principal Data Scientist

Condati

📅 Nov 2021 – Dec 2023   📍 Menlo Park, California (Remote)

- Led the enhancement of a Machine Learning solution for optimizing bidding strategies in digital marketing campaigns, significantly improving financial returns to meet client targets within a year.
- Diagnosed and resolved performance issues in the existing model, leading to a substantial increase in efficiency and effectiveness.
- Collaborated with cross-functional teams to ensure seamless integration and deployment of the improved solution.
- Tech stack: MySQL, Julia, Python, Tensorflow.jl, Torch.jl, MLJ.jl, Flux.jl, Scipy, Numpy, Pandas, Pytorch, Jupyter Notebooks, AWS Sagemaker

AI Researcher & Project Leader - ML for Health

Samsung Research Brazil

📅 Mar 2020 – Nov 2021   📍 Campinas, Brazil (Remote)

- Developed an ML-powered solution for real-time health monitoring on Samsung Galaxy Watch 4.
- Designed data collection protocols and researched AI techniques for health monitoring.
- Prototyped and tested ML solutions for high accuracy and reliability.
- Implemented a memory-constrained solution in C for deployment.
- Led a multidisciplinary team, integrating the app into the Galaxy Watch line globally.
- Tech stack: Python, Julia, C/C++, TensorFlow & Keras, PyTorch & PyTorch Lightning, Pandas, SciPy, NumPy, Scikit-learn, Matplotlib, MLJ.jl, Flux.jl, Jupyter Notebooks, AWS Sagemaker, C, ONNX, custom Python-to-C transpilers, Samsung's Tizen OS

Data Scientist

Poatek IT Consulting

📅 Jun 2019 – Mar 2020   📍 Porto Alegre, Brazil

- Solved combinatorial optimization problems in vehicle routing.
- Applied computer vision and NLP for document cataloging and data analysis.
- Conducted NER and sentiment analysis for actionable insights.
- Developed credit risk models for financial predictions.
- Analyzed geospatial data for decision-making.

PhD in Computer Science

Federal University of Rio Grande do Sul (UFRGS)

📅 Aug 2015 – Jul 2019   📍 Porto Alegre, Brazil

- **AI Ethics**: On Quantifying and Understanding the Role of Ethics in AI Research (GCAI 2018), Assessing Gender Bias in Machine Translation (NCA 2020)
- **Graph Deep Learning & Combinatorial Optimization**: Learning to Solve NP-Complete Problems (AAAI 2019), Graph neural networks meet neural-symbolic computing (IJCAI 2021)

Figures created by me. Resume was created using AltaCV class by LianTze Lim (liantze@gmail.com).