

Cainã Max Couto da Silva

Senior Data Scientist & AI Engineer

Madison - WI, USA

+1 (608) 395 8543 | ✉ cmcouto.silva@gmail.com | [in cmcouto-silva](https://www.linkedin.com/in/cmcouto-silva) | [g cmcouto-silva](https://github.com/cmcouto-silva)

CURRENT EXPERIENCE

Postdoctoral Data Scientist

University of Wisconsin-Madison

Aug 2024 - Current

Madison - WI, USA

- Currently developing a RAG-based **text-to-SQL AI agent** that allows users to interact with the database using Slack.
- Developed [SQLDepts](#): an **open-source Python package leveraging LLMs** to automatically extract table and column dependencies and outputs from complex SQL scripts **100X faster and >300X cheaper** than human expert labor.
- Developed and optimized **cattle mapping** using cutting-edge deep learning models on high-resolution satellite imagery, providing actionable intelligence to combat Amazon deforestation.
- Engineered a pioneering **machine learning classifier** to assess data quality from farm properties at scale, implementing feature engineering on geospatial and entity data that **enabled scalable data integrity checks for the first time**.
- **Lead UW-Madison students** in research projects, mentoring on machine learning and computer vision techniques.
- **Automated database workflows**, reducing manual intervention and boosting team productivity.
- Delivered private **data analysis reports** for stakeholders in Brazil and the USA.

INDUSTRY EXPERIENCE

Data Science Specialist

Schlumberger | World's largest offshore drilling company

Jan 2023 - Jul 2024

Remote - USA

- Deployed multiple machine learning models to predict asset failures and optimize maintenance schedules:
 - Designed **explainable risk indices** using regression and clustering
 - **Published a statistically based model** for detecting failures (see [paper](#))
- Trained models on Azure **Document Intelligence** to extract target components from vendor documents.
- Led an idea for innovation using AI that ranked in the top **12 out of 237** teams worldwide and pitched it to the CEOs.

Data Scientist

AB InBev | World's largest beer brewer company

May 2022 - Dec 2022

Remote - Brazil

- Advanced **forecasting** modeling for multiple product categories with hierarchical reconciliation.
- Built a **scalable data pipeline** for automating pricing and promotion policies on big data using PySpark.

Data Scientist

Remessa Online | Fintech

Oct 2021 - Apr 2022

Remote - Brazil

- Exploratory data analysis, **time series forecasting**, and modeling **customer retention**.
- Engaged with **CRM** and **marketing** staff to present easy-to-understand statistical reports.

ACADEMIC EXPERIENCE

M.B.A in Data Science & Analytics

Universidade de São Paulo

May 2021 - Aug 2023

São Paulo - SP, Brazil

- **Grade: 10** | **Dissertation:** A hybrid approach for predicting and preventing churn [[code](#)]

Ph.D. in Bioinformatics

Universidade de São Paulo

Jul 2016 - Apr 2021

São Paulo - SP, Brazil

- **Grade: 10** | **Thesis:** Identifying natural selection in Native American populations [[code](#)]

CONSULTING EXPERIENCE

Data Science Writer

Provided services for [TrainInData](#)

Apr 2024 - Current

Remote - Germany

- Write SEO-optimized articles on data science [[link](#)]
- Contributed to User Guide pages for [feature-engine](#) Python package.

Data Science Professor

Provided services for DNC Group (Brazilian EdTech)

Oct 2021 - Mar 2025

Remote - Brazil

- Recorded classes about data wrangling, statistics, recommender systems, clustering, model deployment, etc.
- Crafted materials from data wrangling to model deployment.

TECHNICAL SKILLS

Programming: Python, R, SQL, and bash scripting.

Data analysis & statistics: pandas, numpy, scipy, matplotlib, seaborn, plotly, geospatial libraries, etc.

Machine Learning: scikit-learn and related ML packages, feature engineering, model optimization, explainable AI.

Deep Learning: PyTorch, computer vision models, document intelligence, etc.

Generative AI: prompt engineering, RAG, LLM Python frameworks, and LLM providers.

Cloud computing & Big Data: AWS, GCP, SQL, MySQL, Postgres, SparkSQL (PySpark), and BigQuery.

MLOps and model deployment: Git, GitHub/GitLab, GitHub Actions, CI/CD, Docker, FastAPI, MLFlow, W&B.

Data Platforms: Databricks, Dataiku.

PUBLICATIONS

- **Couto-Silva, C. M.**, et al. (2024). Mitigating Nonproductive Time: A Novel Algorithm for Dsl Fault Detection. *OnePetro*. DOI: [10.2523/IPTC-24515-MS](https://doi.org/10.2523/IPTC-24515-MS).
- **Couto-Silva, C. M.**, et al. (2023). Indigenous people from Amazon show genetic signatures of pathogen-driven selection. *Science Advances*, 9(10). DOI: [10.1126/sciadv.abo0234](https://doi.org/10.1126/sciadv.abo0234).
- Castro e Silva, M. A., Ferraz, T., **Couto-Silva, C. M.**, et al.. (2021). Population Histories and genomic diversity of South American natives. *Molecular Biology and Evolution*, 39(1). DOI: [10.1093/molbev/msab339](https://doi.org/10.1093/molbev/msab339).
- Jacovas, V. C., **Couto-Silva, C. M.**, et al. (2018). Selection scan reveals three new loci related to high altitude adaptation in native Andeans. *Scientific Reports*, 8(1). DOI: [10.1038/s41598-018-31100-6](https://doi.org/10.1038/s41598-018-31100-6).

TALKS

Scipy 2025

Building an AI Agent for Natural Language to SQL Query Execution on Live Databases

Jul 2025 (scheduled)

Tacoma - WA, USA

- 4-hours tutorial | Live coding

PyData Global 2024

PyTorch Workflow Mastery: A Guide to Track and Optimize Model Performance

Dec 2024

Remote, Global

- 90-minute tutorial | Live coding | [Code Repository](#)

Nubank Meetup 2024

Data Science Applied to Multiple Sectors: Challenges, Solutions, and Lessons Learned

Oct 2024

Remote, Brazil

PyData Global 2023

Introduction to ML Pipelines: How to Prevent Data Leakage and Build Efficient Workflows

Dec 2023

Remote, Global

- 2-hours workshop | [Code repository](#)

ADDITIONAL INFORMATION

LANGUAGES

- Portuguese (native) - English (professional / C1) - Spanish (intermediate)

AWARDS

- **Best academic performance** award (2013)
- Highlighted didactic [web app](#) for clustering on **Streamlit** [Weekly Roundup](#)

OPEN-SOURCE CONTRIBUTION

- Contributed to [Feature-engine](#), an open-source Python package implementing various feature engineering techniques.
- Developed [ProjectLens](#), an open-source Python package that generates a well-structured text file with comprehensive project snapshots, optimizing context for LLM analysis and code review.

INTERNATIONAL EXPERIENCE

- Postdoctoral researcher position in Wisconsin, USA.
- PhD's internship in Barcelona - Spain.
- Master's internship in Buenos Aires - Argentina.
- Attended multiple conferences in South America, the USA, and Spain.