Bruna Zamith Santos, PhD Candidate

bzamith.github.io

Email : HIDDEN
Mobile : HIDDEN

Main Areas of Interest

- **Data Science, Machine Learning, Research & Development, Software Engineering**

Professional Experience

- **Amazon** — Remote - Brazil
  - Applied Scientist II — Feb. 2024 – Present
  - Data Scientist II — Mar. 2022 – Jan. 2024
    - Own end-to-end supervised ML for Catalog and Supply Chain: business framing, data/feature pipelines, model training, offline and online evaluation, deployment, monitoring, and MLOps.
    - Delivered models credited with $50M+ incremental revenue and enabling 1.3M+ new catalog product launches through scalable decisioning services and data products.
    - Partnered with PM/Eng/Ops to prioritize a multi-quarter roadmap, negotiate trade-offs (latency/cost/quality), and unblock adoption across LatAm businesses.
    - Program Owner, GenAI capacitation: designed curriculum and delivered trainings and office hours for different job roles across 12 marketplaces (10k+ employees); measured adoption and learning outcomes.
    - Published 2 internal Amazon papers (applied ML/GenAI); evangelized best practices in experimentation, evaluation, and productionization.
    - **Technologies:** AWS, Python, SQL, Typescript
    - **Domains:** GenAI, Forecasting, Classification, A/B Testing, MLOps

- **Amazon** — Sao Paulo, SP - Brazil
  - Software Development Engineer II — Dec. 2021 – Feb. 2022
  - Software Development Engineer I — Jul. 2020 – Nov. 2021
  - Software Development Intern — Jan. 2020 – Jul. 2020
    - Built and operated large-scale front-/back-end services integrating with Amazon core systems to improve seller & buyer experiences.
    - Led design reviews, API contracts, performance testing, reliability, and on-call for high-throughput workloads.
    - Shipped resilient, well-instrumented services adopted by core retail flows in Brazil.
    - **Technologies:** AWS, Java, Python, Typescript
    - **Domains:** Distributed Systems, APIs, Observability, Reliability

- **Serasa Experian** — Sao Carlos, SP - Brazil
  - MIS Intern — Aug. 2019 – Dec. 2019
    - Delivered Finance dashboards and automated data pipelines; improved reporting accuracy and turnaround time by 40%.
    - **Technologies:** SQL, SAS, Tableau
    - **Domains:** Business Intelligence, Data Engineering

Education & Research

- **Federal University of Pernambuco (UFPE)** — Remote — Brazil
  - PhD in Artificial Intelligence — Feb. 2025 – Present
    - **Topic:** Explainable AI (XAI) for Time Series Classification.
    - **Technologies:** Python (PyTorch/NumPy/Pandas), LaTeX
    - **Domains:** XAI, Time Series, Instance Hardness
    - **GPA:** 10/10

- **Federal University of Sao Carlos (UFSCar)** — Remote - Brazil
  - Masters in Artificial Intelligence — Aug. 2021 – Dec. 2023
    - **Project:** Climate Variables Forecasting & Forest Fire Risk Classification in the Brazilian Pantanal. Pantanal faces recurrent fires driven by climate; existing indexes are biome-agnostic, limited in variables, and lack multi-day predictions.
    - **Approach:** Built a two-stage ML system: (1) forecast climate variables; (2) classify daily fire-risk levels using predicted covariates. Compared exhaustive hyperparameter search to Genetic Algorithms (GA) for multi-objective optimization.
    - **Results:** System outperformed standard fire-risk indexes on key classes and enabled forward-looking risk.
    - **Technologies:** Python, R
    - **Domains:** Time Series, Forecasting, Classification, Genetic Algorithms
    - **GPA:** 10/10

- **Katholieke Universiteit Leuven (KU Leuven)** Kortrijk, Flanders - Belgium
  Research Intern Sep. 2017 – Dec. 2017
  - **Project:** Predicting Protein Functions via Interaction Prediction. Funded by Sao Paulo Research Foundation (FAPESP).
  - **Approach:** Modeled protein function prediction as Hierarchical Multi-label Classification (HMC) over interaction networks; implemented data processing and evaluation pipelines.
  - **Outcome:** Delivered reproducible code and benchmarks; informed subsequent HMC research at BioMaL/KU Leuven.
  - **Technologies:** Java
  - **Domains:** HMC, Bioinformatics
- **Federal University of Sao Carlos (UFSCar)** Sao Carlos, SP - Brazil
  Bachelor in Computer Engineering Mar. 2015 – Jul. 2020
  - **GPA:** 8.75/10

Programming Skills

- **Languages:** Python (6 years); Java (5 years); SQL (3 years); Typescript (2 years); R (2 years); C++ (6 months)
- **Others:** AWS, Git, Linux, Latex

Language

- **Portuguese:** Native
- **English:** Fluent (Cambridge English Certificate ESOL Intl)

Presentations, Proceedings, and Papers

- **"A New Time Series Framework for Forest Fire Risk Forecasting and Classification"** Zamith B., Soriano, B., Narciso, M., Furtado, D., Cerri R. (2023). International Joint Conference on Neural Networks (IJCNN).
- **"Predictive Bi-Clustering Trees for Hierachical Multi-label Classification"** Zamith B., Nakano, K. F., Cerri R., Vens C. (2020). European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD).
- **"Strategies for Selection of Positive and Negative Instances in the Hierarchical Classification of Transposable Elements"** Zamith B., Pereira, T. G., Nakano, K. F., Cerri R. (2018). Brazilian Conference on Intelligent Systems (BRACIS).
- **"A Genetic Algorithm for Transposable Elements Hierarchical Classification Rule Induction"** Pereira, T. G., Zamith B., Cerri R. (2018). IEEE Congress on Evolutionary Computation (IEEE CEC).
- **"A New Machine Learning Dataset for Hierarchical Classification of Transposable Elements"** Zamith B., Cerri R. (2016). National Meeting of Artificial and Computational Intelligence (ENIAC).
- **"Decisions Trees for Hierarchical Classification of Transposable Elements"** Zamith B., Gomes Mantovani R., Schietgat L., Vens C., Cerri R. (2016). Proceedings of the 25th Belgian-Dutch Machine Learning Conference (Benelearn).

Certifications

- **Fundamentals of Probability and Statistics for Data Science (2025):** 30 hours online course provided by University of Sao Paulo (USP).
- **Practical Multi AI Agents and Advanced Use Cases with crewAI (2025):** 3 hours online course provided by DeepLearning.ai and crewAI.
- **Practical Time Series Analysis (2024):** 25 hours online course provided by The State University of New York (SUNY).
- **Generative AI with Large Language Models (2024):** 16 hours online course provided by DeepLearning.ai and Amazon Web Services (AWS).
- **Sequence Models (2022):** 38 hours online course provided by DeepLearning.ai.
- **Sequences, Time Series and Prediction (2022):** 23 hours online course provided by DeepLearning.ai.
- **Introduction to AWS (2020):** 5 hours online course provided by A Cloud Guru.
- **Version Control With Git (2019):** 12 hours online course provided by Atlassian.
- **Software Development Processes and Methodologies (2019):** 18 hours online course provided by University of Minnesota.

Other Projects

- **School of AI - Health Hackathon (2019):** Team project aimed to develop an app which seeks to reduce patients waiting time (at hospitals queue). The patient types what he is feeling and then via Natural Language Processing the symptoms are classified into a medical specialty.
- **Intel IoT Roadshow (2015):** Team project aimed to develop a "smart" bathroom making use of relevant capacity utilization data, and trigger "on demand" cleaning process alert, which leads to resources use optimization.