

RAFAEL MOREIRA DA COSTA NUNES RIBEIRO

DATA SCIENTIST & ML ENGINEER

📍 Braga, Portugal

☎ (+351) 916 544 008

🌐 [linkedin.com/in/ribeiorafa](https://www.linkedin.com/in/ribeiorafa)

🐙 github.com/rmcnribeiro

✉ rmcn.ribeiro@gmail.com

HIGHER EDUCATION

Bachelor's Degree in Electronic and Computer Engineering (GPA 7.6/10)

📅 Mar 2018 - Sep 2025

Universidade Federal do Rio de Janeiro (UFRJ)

[Thesis Title: "10-Year Dementia Prognosis: Applying Machine Learning Models and Explainable AI Techniques to Identify Modifiable Risk Factors"](#)

Academic Exchange Program

📅 Sep 2022 - Dec 2022

Aalto University - School of Electrical Engineering

EXPERIENCE

Casais Group (Braga, Portugal)

Casais Group is one of the largest construction companies in Portugal and stands out for its continuous focus on innovation and sustainable construction. In 2026, the company is releasing "GoTwin", a **Digital Twins** and **Facility Management software** with **real-time building visualization**, integrated asset and **maintenance management**, and **data-driven insights** that optimize operations throughout the building lifecycle.

Internet of Things (IoT) Engineering Trainee (*GoTwin Project*)

📅 Jun 2025 - Present

- Designed and implemented a gateway for the IoT system using **NodeJS**, **WebSockets** and **MongoDB**
- Contributed to the architecture, modelling and implementation of the **system's backend** IoT layer
- Implemented a **VPN (Wireguard)** for security, gateway maintenance and remote access (**SSH** and **SFTP**)
- Administered Linux environments, created **logging**, **monitoring** and developed **automation** flows using Shell Script
- Set up, configured and maintained **cloud infrastructure** on **AWS**
- Managed **containers** and **virtual machines (VMs)** on the gateway using **Docker** and **Proxmox**
- Maintained, monitored and improved the system's **RestAPI** (NodeJS)

TYR Energia (Rio de Janeiro, Brazil)

Tyr Energia is a company operating in the free energy market, providing sustainable and cost-efficient solutions with a strong focus on renewable energy and intelligent monitoring for businesses across multiple sectors.

Data Intelligence Analyst

📅 Nov 2024 - Apr 2025

- **Automated** the company's billing process using **Python** and **third-party APIs**
- Deployed a production LLM-based retrieval-augmented generation (RAG) system using Python and LangChain
- Developed web **automation** bots using Selenium (Python)
- Designed, implemented and maintained the company's **cloud infrastructure** on Oracle Cloud

Data Intelligence Intern

📅 Sep 2023 - Nov 2024

- Designed and implemented the **MySQL database** for the company's commercial department
- Developed **automation pipelines** for multiple business areas
- Built automated **dashboards** using Looker Studio and Power BI for the Finance and Commercial teams
- **Automated** the commercial team's **ETL data flow** using **Python** and **SQL** on Oracle Cloud, reducing the data refresh interval from weekly to 15 minutes
- Defined and implemented **KPIs** for the Finance and Commercial teams

UFRJ Analytica (Rio de Janeiro, Brazil)

📅 Oct 2019 - Aug 2022

UFRJ Analytica is the data science competition team of the Federal University of Rio de Janeiro.

Co-founder and Head of Competition

- Led teams in **hackathons** and **datathons**, including Hacking.Rio 2021, where two Analytica teams won among 280+ competing teams
- Managed a group of 20+ members, coordinating the development of **educational materials** and study plans for new joiners
- Oversaw the recruitment process for new members and supported the acquisition of new projects and partnerships
- Collaborated on **data projects** with the **Rio de Janeiro City Hall** and the Rio de Janeiro State Health Department

Data Science for Smart Cities - Extension Group (Rio de Janeiro, Brazil)

📅 May 2020 - Apr 2022

Project coordinated by Heudson Mirandola (IM/UFRJ) with the goal of preparing students to work in the field of Data Science, focusing on projects with social impact.

Extension Group Member - IM/UFRJ

- Performed data mining, exploratory **data analysis** and visualization to identify patterns and insights
- Researched and applied techniques for handling **missing data**, **imputation**, and **variable selection**
- Built and evaluated **machine learning models**, including **time series** analysis and forecasting

Undergraduate Technological Development and Innovation Researcher (Machine Learning)

- **Researched** and applied state-of-the-art **document classification** techniques for large-scale **fraud detection**, achieving over 95% accuracy and a high F2 score
- Collaborated with the CVM **Data Science team** and UFRJ Mathematics Institute faculty to design, validate, and refine models
- Built an **NLP pipeline** in Python (TF-IDF, spaCy, NLTK, Word2Vec), including feature selection using Mutual Information and hyperparameter tuning with GridSearch
- **Deployed** the best-performing model for internal use via a Streamlit application

VOLUNTARY WORK

COVID-19 Data Analysis and Prediction

Extension Volunteer - GT Coronavírus/UFRJ/Rio de Janeiro City Hall

- Provided **data-driven decision-making insights** for the Rio de Janeiro City Hall during the pandemic
- Collected, processed, analysed, and created **data** visualizations for COVID-19 data from various countries
- Developed **predictive models** for new cases, deaths, and hospitalizations using Python
- **Implemented a compartmental model** based on the latest epidemiological literature, used in the Public Prosecutor's Office of the State of Rio de Janeiro (MPRJ)'s decision-making process regarding lockdown in the city

COVID-19 Dashboard

Extension Volunteer - GT Coronavírus/UFRJ/Rio de Janeiro State Department of Health

- **Processed, cleaned, and aggregated COVID-related data** for the Rio de Janeiro State Health Department
- Developed and maintained a data visualization **dashboard** for the coronavirus pandemic in the state of Rio de Janeiro
- Created and maintained the "Covidímetro", an indicator of the contagion rate per epidemiological week

HACKATHONS

3rd Place (Team Charneira) - Hacktudo: 2024

Hackathon promoted by Bemobi. 107 teams subscribed to the competition.

- Topic: How can Bemobi use **GenAI** to provide a better experience for its customers?
- Solution developed: A **B2B2C ecosystem** called "OmniChat" - a **chatbot** with contextual understanding for customers, and for decision-makers, a dashboard featuring each conversation summarized by **GenAI**, along with suggestions for new products based on this history.

Winner Team (UFRJ Analytica) - Hacking.Rio 2021

Annual hackathon, with "Vale" as main sponsor. This year, 282 teams registered for the event.

- Development of an **interactive dashboard** to facilitate the visualization of climate action indicators
- Provided information to guide Vale's decision-making, aiming to minimize incidents in conservation units

Winner Team (UFRJ Analytica) - Open Data Day (ODD) 2021

Datathon organized by the Base dos Dados organization in partnership with Fundação Getúlio Vargas (FGV).

- Creation of a document focused on understanding how the use of **open data** can help reduce inequalities in Brazil, based on the **Sustainable Development Goals (SDGs)**. The submission can be viewed at this [link](#)

Winner - Master the Mainframe 2019 Part 2 (IBM)

Master the Mainframe contest is a Mainframe programming challenge that is organized annually by IBM.

- Challenges in the **IBM Z Mainframe** environment, including creating and manipulating data partitions, modifying programs, as well as creating and executing JCLs
- Certificates from the [2019](#) and [2020](#) editions can be viewed via the links and on LinkedIn

LANGUAGE SKILLS

Portuguese

Native (C2)

English

Fluent (C1) (TOEFL IBT 110/120)

Spanish

Intermediate (B1)

Finnish

Beginner (A1)

COURSES

- Datacamp - Introduction to NLP with Python
- Huawei - Artificial Intelligence and Computer Vision
- Huawei - Data Storage
- Huawei - Internet of Things (IoT)
- Huawei - FTTX Networks (Fiber-to-the-X)
- Huawei - Data Communication and Wireless Networks

HARD SKILLS

Python (Libraries and Models)

Machine Learning and GenAI: Sklearn | XGBoost | CatBoost | Facebook Prophet | Pandas | Numpy | SciPy | LangChain | PineconeAPI | OpenAI API | Gemini API | SHAP
Data Visualization: Matplotlib | Seaborn | Plotly | Streamlit | Dash |
Others: Requests | Flask | FastAPI | Selenium | BeautifulSoup | SQLAlchemy

System Infrastructure

Databases: MySQL | PostgreSQL | MongoDB | Firebase | Pinecone
Languages: JavaScript | C | C++ | Shell Script | Powershell
Infrastructure: Docker | AWS | Oracle Cloud | Cron | Nginx | CI/CD | Linux | Proxmox | Git/GitHub | Wireguard VPN | SSH | SFTP

Engineering

LTSpice | OrCad | KiCAD | AutoCAD | Ultimaker Cura | ESP32 | Arduino | Node-RED

Others

PowerBI | Looker Studio | LaTeX | Excel | Canva | Visio | MATLAB | n8n | Claude Code