ATHOS MEKANNA MORAES

As a Data Scientist with a background in Physics, I've been applying data analysis and machine learning across various sectors including retail, supply chain, and healthcare. My professional journey includes consultancy and self-employment, leveraging diverse Data Science skills from Deep Learning to Data Engineering, and implementing APIs and user interfaces. Having recently completed a Master's in Bioinformatics, I'm enthusiastic about exploring the intersection of data science and healthcare. Through my thesis, I've gained experience in trending areas related to Generative AI and deepened my understanding of how data science can be practically applied to solve biological challenges.

CONTACT

✉ athos.m.moraes@gmail.com
📞 +351 928060194
📍 Porto - Portugal
in Athos Moraes

TECHNICAL SKILLS

Data Science

Python ●●●●●
PySpark ●●●●○
SQL ●●●●○
Data Processing ●●●●●
(Pandas, Numpy, Spark DF...)
Machine Learning ●●●●○
(Spark ML, SkLearn, ...)
Deep Learning ●●●○○
(TensorFlow, Keras, PyTorch)
Statistics ●●●●●
(Numpy, Scipy, ...)

Operating System/Data Storage

Linux ●●●●○
AWS ●●●○○
(EC2, S3, Athena, SageMaker)
GIT ●●●●○
(GitHub, Bitbucket, ...)

User Experience

Python ●●●●○
(Dash, Streamlit, Plotly, Matplotlib, ...)
Front-End ●●●○○
(Django, Angular, React, Javascript, ...)
Back-End ●●●○○
(Flask, FastAPI, Django)

Languages

Portuguese ●●●●●
English (C1) ●●●●○

🎓 EDUCATION

2021 - 2024
📍 University of Porto - UP
Master's degree, Bioinformatics

2013 - 2017
📍 University of São Paulo - USP
Bachelor's degree, Physics

⚙ WORK EXPERIENCE

12/2020 - 01/2024
📍 Yhub, São Paulo
Data Scientist
In my current role as a Data Science Consultant and Developer, I specialize in creating data-driven solutions using Python and PySpark for a diverse range of challenges within the retail industry. My work encompasses Natural Language Processing, classification algorithms, product similarity determination through vectorization and clustering, data normalization, and demand forecasting. I also focus on implementing big data handling solutions, deep learning techniques for product attribute recognition, and automating report generation. Throughout this position, I have gained extensive experience working with large-scale datasets and using AWS tools to optimize data processing and analysis.

03/2019 - 11/2020
📍 CTI Global, São Paulo
Data Scientist
As a Data Science Consultant, I collaborated with prominent clients to address complex challenges within the Supply Chain domain. During this role, I was responsible for creating demand prediction algorithms, determining optimal safety stock levels, and applying discrete event simulation alongside machine learning techniques to create comprehensive supply chain models. This experience not only allowed me to enhance my data science expertise but also develop skills in data engineering, good coding practices, deployment pipeline, and maintenance of solutions.

10/2018 - 01/2019
📍 WeMind, São Paulo
Front-end developer
As a Web Developer with a primary focus on front-end development, I contributed to the creation of secure and robust web environments for diverse clients. My work centered around integrating data science solutions into these web applications, often in the form of interactive dashboards, allowing clients to better understand and utilize the data-driven insights.

01/2018 - 08/2018
📍 E/OU MRM, São Paulo
Data Scientist
As a Data Scientist and Business Intelligence Analyst, I focused on key business objectives such as reducing customer churn and optimizing customer segmentation. By implementing classification models for churn prediction and utilizing clustering algorithms for segmentation, I provided actionable insights to drive retention campaigns. Additionally, I maintained and improved data visualization dashboards to effectively communicate results, supporting collaboration and data-driven decision-making among internal teams.

⚙ ACADEMIC EXPERIENCE

📅 Oct 2022 - Present (8 months)
📍 Universidade do Porto

Master's Student

In my master's thesis in bioinformatics and computational biology, I developed a system capable of generating automatic reports from lung biopsies based on digital images of these samples. From a computational perspective, the project primarily involved the fields of Computer Vision and NLP, with a particular focus on the use of Large Language Models (LLMs) and Vision Language Models (VLMs) for image processing and report generation. Clinically, the work required knowledge in histopathology, both for the design of the system and for evaluating its performance. Through this experience, I have become capable of understanding the most current technologies in the realm of Generative AI in mathematical detail and using them effectively to solve a problem in the medical field.

📅 2015 - 2017
📍 University of São Paulo

CNPq researcher

During my bachelor's as a CNPq researcher at the Institute of Physics of the University of São Paulo (IF-USP), I worked on a research project focusing on the atomic-level magnetic structures in nanomaterials. My work revolved around employing the Density Functional Theory (DFT) to understand these structures better. Central to this project was the application of Local Spin Density Approximation (LSDA) programmed in Fortran to solve Schrödinger's equation for nanosystems. All computations were handled within a Linux environment to ensure the overall efficiency of the simulations. The subsequent analysis of the results was conducted using Python, which allowed for a understanding of the magnetic properties of the studied nanomaterials.

📅 2012 - 2013
📍 University of São Paulo

CNPq researcher

As a CNPq researcher at the Institute of Astronomy, Geophysics and Atmospheric Sciences of the University of São Paulo (IAG-USP), I played a role in a project focusing on understanding the monthly mean radiation balance on the surface of the Antarctic region around the Brazilian base *Comandante Ferraz*. The project necessitated processing and analyzing a vast volume of data collected from a radiation intensity sensor installed at the base. Leveraging descriptive statistical methods for analysis, we aimed to shed light on the net radiation characteristics in the Antarctic region. This experience not only improved my data analysis skills but also offered a practical perspective on applying these techniques in a real-world research setting.

📖 PUBLICATIONS

📅 2024
📍 Springer - Machine Learning
🔗 Integration of multi-modal datasets to estimate human aging

Developed machine learning models to predict biological age using a combination of genetic, epigenetic, and histological data. Key techniques included tissue-specific modeling and advanced image analysis, resulting in highly accurate age predictions, particularly using DNA methylation data. The study highlighted the effectiveness of integrating multiple data types for improving age prediction models, demonstrating potential advancements in understanding the biology of aging.

📅 2024
📍 bioRxiv
🔗 The molecular impact of cigarette smoking resembles aging across tissues

Conducted a multi-tissue, multi-omic analysis to explore the molecular mechanisms by which smoking accelerates aging. This study identified that smoking induces gene expression changes, DNA methylation alterations, and tissue damage, many of which overlap with aging processes. Key findings include smoking-induced inflammation, hypomethylation at enhancers, and persistent epigenetic changes that mirror aging effects. These results suggest that smoking may directly contribute to accelerated aging through specific molecular pathways, providing new insights into the long-term health impacts of tobacco use.

📅 2023
📍 Springer - Discovery Science
🔗 Predicting Age from Human Lung Tissue Through Multi-modal Data Integration

Developed age prediction models using multi-modal datasets (transcriptomics, methylation, histological images) from 793 human lung tissue samples to promote healthy aging strategies. Overcame technical challenges such as missing data and high feature dimensionality by employing gradient boosting tree and convolutional neural network models. Improved model performance through data-centric approaches, addressing covariates and skewed age distribution. Achieved a median absolute error of 4 years in age prediction by combining models into a multi-modal ensemble. Provided insights into multi-modal data integration and imbalanced data prediction.