Marcos de Souza Oliveira

Coimbra, Portugal
marcosd3souza@gmail.com
https://marcosd3souza.github.io/

Preamble

I am a highly skilled researcher and professional with over 14 years of experience in software development, Artificial Intelligence (AI), Machine Learning (ML), and data analytics. Throughout my career, I have led and contributed to a wide range of projects with international collaborations. My work spans academic research, industrial innovation, and educational leadership, with a particular focus on data analysis. My main interest is to develop meaningful, cost-effective, and explainable AI models through exploratory data analysis and optimization. Specifically, I focus on deep clustering, deep graph clustering, evolutionary algorithms, and dimensionality reduction methods. I am committed to advancing intelligent systems that are not only technically robust but also interpretable and impactful in real-world applications.

Education

2025 **Ph.D. in Computer Science**, *Federal University of Pernambuco*, Recife, Brazil
Research Area: Machine Learning
Title: *Unsupervised Feature Selection and Deep Subspace Clustering for Exploratory High-Dimensional Cluster Analysis*
This research addresses the challenges of analyzing high-dimensional datasets with limited samples by proposing two novel strategies: an unsupervised feature selection method and a subspace clustering approach. Unlike traditional dimensionality reduction techniques, the proposed methods preserve interpretability and uncover latent structures within specific feature subsets. Experimental results on synthetic and real-world datasets demonstrate superior performance compared to state-of-the-art techniques.
URL: https://repositorio.ufpe.br/handle/123456789/62451

2018 **Master in Computer Science**, *Federal University of Pernambuco*, Recife, Brazil
Research Area: Machine Learning
Title: *Unsupervised Feature Selection Methodology for Clustering in High-Dimensional Datasets*
This work proposes a methodology to evaluate and combine unsupervised feature selection methods, addressing key questions on parameter tuning, feature ranking, and method fusion. The approach was validated on datasets from domains such as image processing and bioinformatics, showing improvements in clustering performance metrics like accuracy, NMI, and Corrected Rand Index.
URL: https://repositorio.ufpe.br/handle/123456789/33642

2016 **Specialization in Software Engineering**, *University of Pernambuco*, Recife, Brazil
Research Area: Software Engineering
Title: *SOA-Based Architecture Proposal for Avatar Educação*
This work proposes a service-oriented architecture to decouple the Avatar Educação virtual assistant from the Moodle platform, enhancing interoperability and scalability in e-learning environments. The architecture aims to improve user interaction and facilitate integration with diverse educational systems.
URL: http://revistas.poli.br/index.php/repa/article/view/528

2014 **Bachelor in Software Engineering**, *Faculdade dos Guararapes*, Recife, Brazil
Research Area: Software Engineering
Title: *EDUCAPP: Hybrid Educational Support Application for Primary Education*
This project presents the development of a hybrid mobile application using HTML5 and Cordova to support academic performance in Brazilian primary schools. It emphasizes accessibility and usability, integrating native Android features with web technologies to deliver an inclusive educational tool.
URL: http://dx.doi.org/10.6084/m9.figshare.30383980

Professional Experience

2025–Present **Postdoctoral Researcher**, *CISUC, University of Coimbra*, Coimbra, Portugal
At CISUC, I am engaged in cutting-edge research integrating Machine Learning, Digital Twins, and Augmented Reality to enhance quality inspection in industrial production lines. My focus is on developing interpretable AI systems using unsupervised learning techniques to support expert decision-making. This role allows me to contribute to the modernization of manufacturing processes through intelligent, human-centered technologies.

2023–2025 **Senior Machine Learning Research Engineer**, *Instituto Nacional de Telecomunicações (Inatel)*, Santa Rita do Sapucaí, Brazil
At Inatel's Competence Center (ICC/PDI), I lead R&D projects focused on AI/ML for industrial applications. My responsibilities include designing statistical models, deploying scalable AI solutions using MLOps best practices, and conducting studies on frameworks and platforms. I prepare functional documentation, write pre-sale proposals, and publish in Scopus-indexed journals. I also support QA, design, and development teams, review code for clean architecture, and foster knowledge sharing through technical presentations. My work spans technologies such as Network as a Service (NaaS), Edge Computing, and IoT, positioning me as a technical leader in industrial AI innovation.

2019–2023 **Data Scientist**, *CESAR Innovation Center*, Recife, Brazil
As a Data Scientist at CESAR, I led initiatives across AI, IoT, and software engineering. In 2023, I designed scalable, secure data solutions for international clients and developed an internal tool using LLMs to automate technical report generation. In 2022, I built a probabilistic model to predict employee turnover, supporting strategic decisions at the executive level. During the Dell AI Platform project, I contributed to backlog planning, anomaly detection in time series, and real-time resource allocation models—culminating in a publication in a top-tier journal. Between 2019 and 2021, I developed route optimization systems and queue management models for Samsung's customer service platform, applying operational research to enhance service efficiency.

2017–2018 **Software Engineer**, *CESAR Innovation Center*, Recife, Brazil
In 2018, I contributed to the Qless System project, designing low-cost ETL pipelines using AWS components (e.g., Lambdas and SQS) to support predictive APIs. I built a queue management simulator for ML preprocessing and developed both frontend (AngularJS) and backend (NodeJS) components within Dockerized EC2 environments. I also implemented a natural language classifier to detect negative sentiment in customer feedback. In 2017, I developed Android PoCs including a noise reduction engine for voice audio and a multimedia gallery app that used Spotify's API for music recognition. These projects showcased my ability to merge AI with user-centric mobile experiences.

2016–2017 **Android Engineer**, *Stefanini*, Recife, Brazil

At Stefanini's Document Solutions division, I led the development of a mobile application for digital bank account creation, which gained traction with major financial institutions like Itaú. I integrated C++ image segmentation engines into Android apps via JNI and delivered AAR-format engines for BBVA Argentina, collaborating directly with their tech leadership. I also ensured the portability of DSPhoto, a bank check recognition engine, across a wide range of smartphones—from mainstream brands like Samsung and Apple to emerging manufacturers such as Xiaomi. This role combined applied research with client-facing delivery in fintech and mobile imaging.

2015–2016 **Software Engineer**, *Center of Informatics, Federal University of Pernambuco*, Recife, Brazil

As part of the Samsung/Cin-UFPE R&D initiative, I contributed to projects aimed at delivering high-quality software solutions for Samsung Electronics. I developed a backend API for a social network intended for the Rio 2016 Olympics, ensuring compliance with security, scalability, and OAuth integration standards. I also participated in accessibility-focused research, creating guidelines for visually impaired users to improve app usability via Google Talkback. Additionally, I helped build cross-project tools to verify code compliance with Samsung's standards, including checks for English language usage, copyright headers, and performance benchmarks. This experience honed my skills in software quality assurance and inclusive design.

2014–2015 **Java Software Developer**, *MV S/A*, Recife, Brazil

MV is a leading health tech company that supports the operations of major hospitals throughout Brazil. I was part of the supply module team, contributing to both legacy system maintenance and platform migration from Oracle Forms (MV2000) to Java Web (SoulMV). My work included SQL scripting with Oracle 12g, PL/SQL development, and report generation using iReports and JasperReports. I developed Java web applications using Struts, Flex, and the MVC pattern, and facilitated onboarding for new team members. This role provided me with a strong foundation in enterprise healthcare systems and cross-functional development environments.

2011–2014 **Lead Technical Support Analyst**, *Qualyfit Systems and Consulting*, Recife, Brazil

At Qualyfit, a company specializing in software development and consultancy for over 90 gym businesses across Brazil, I led the technical support team and played a pivotal role in systems training, migration, and documentation. My responsibilities spanned across access control, CRM, and financial modules, where I provided both remote and on-site consultancy to clients, advising on strategic metrics such as average ticket and turnover using balanced scorecard methodologies. Technically, I was deeply involved in SQL scripting with SQL Server, MSSQL, and PostgreSQL, database administration including migration and backup policies, and the development of an internal ERP system using the .NET framework. I also mentored teams on KPI design and sales planning strategies. This experience was instrumental in shaping my interest in business management, leading me to pursue an MBA in the field.

Teaching Experience

2021–2023 **Facilitator and Instructor**, *CESAR School*, Recife, Brazil

CESAR School is the educational branch of the CESAR Innovation Institute, offering boot camps, bachelor, master and professional degrees using a problem-based learning (PBL) approach. As a facilitator, I contributed to initiatives such as the TechDesign FAST program, which aims to reintegrate professionals into key areas such as design, testing, and machine learning. As an instructor in the Data Specialization, I teach the Unsupervised Learning discipline, covering dimensionality reduction, clustering algorithms, unsupervised metrics, and deep clustering. I also mentor undergraduate students in software-based projects applying ML and design fundamentals to real-world challenges.

2020 **Professor**, *Centro Universitário Tiradentes (Unit)*, Recife, Brazil

I taught core Computer Science courses, including Algorithms, Programming Logic and Data Structures, Computing and Programming, and Operating Systems Infrastructure. My teaching focused on developing logical reasoning and programming skills, covering topics from algorithm design and data structures to memory management, process scheduling, and system concurrency, providing students with a solid foundation in both software and system-level computing concepts.

2019 **Professor**, *Unibratec*, Recife, Brazil

Courses taught: Object-Oriented Paradigm, Design Patterns, unit tests (JUnit), Android Mobile Development, Java Programming.

Academic Activities

2023 Supervision of student **Anderson Henrique Custodio Bento**, in the *Postgraduate Program in Data Analysis and Engineering (Specialization)*, under the title *"Análise de Clusters de Percepções dos Brasileiros Sobre Meio Ambiente e Preservação da Amazônia"* — defended on May 30, 2023. DOI: 10.6084/m9.figshare.30399622

Research Projects

2024–2026 **E2E Digital Twin** (TRL 4) — Funded by PT2030. This project aims to revolutionize the quality inspection of premium natural cork stoppers by integrating digital and physical processes. It involves the creation of a Digital Twin for each cork unit and its associated inspection, quality control, and assembly stages. Machine learning models are used to reduce classification overlap between adjacent quality grades, while augmented reality devices bridge tacit worker knowledge with digital insights. The initiative also includes multimedia training tools to capture expert inspection strategies and accelerate onboarding of new employees. **My contribution:** Development of unsupervised deep learning models to discover natural patterns and perform data labeling for supervised classification. **REF:** https://vrlab.dei.uc.pt/projects/e2e-digital-twin/

2023–2025 **EDGE5G** (TRL 4) — Funded by EMBRAPII. A cutting-edge Industry 4.0 testbed leveraging private 5G networks to enable high-precision indoor geolocation of industrial assets such as AGVs and AMRs. The project also addresses cybersecurity for IoT terminals and explores edge computing, augmented reality, and AI for productivity gains. **My contribution:** Authored a scientific paper on indoor geolocation using OFDM pilot signals and trained deep learning models for centimeter-level accuracy. **REF:** https://embrapii.org.br/...pesquisa-analisara-impactos-do-5g...

2024–2025 **ICE – Sheet Inspection Tool** (TRL 4) — Funded by EMBRAPII. This project automates the detection of anomalies in pipelines using Gauss data from triaxial Hall effect sensors. It includes real-time monitoring with human-in-the-loop feedback, enabling experts to validate and correct automatic detections. The system spans hardware, firmware, and software components, with mobile and desktop interfaces for field and lab use. **My contribution:** Developed the anomaly detection algorithm and authored the functional requirements documentation. **REF:** http://eticaas.com.br/end-ensaios-nao-destrutivos/

2022–2023 **Dell Real-Time AI Platform** (TRL 7) — Funded by Informatics Law. A resource management platform for Dell's HPC clusters, designed to optimize allocation of virtual machines and applications for data science teams. The project tackled challenges in scalability, observability, and anomaly detection, integrating MLOps practices and predictive modeling. **My contribution:** Designed heuristic and optimization methods for real-time resource allocation. **REF:** https://doi.org/10.1016/j.future.2023.08.014

2017–2021 **Samsung Qless** (TRL 9) — Funded by Informatics Law. An intelligent queue management system deployed in Samsung concept stores, combining AI-based service optimization with real-time KPI dashboards. Evolutionary algorithms (PSO/GA) were used to model queue priorities and predict optimal service sequences. The system significantly improved customer experience and operational efficiency. **My contribution:** Full-stack development of smart queue systems using operational research and evolutionary optimization techniques. **REF:** https://news.samsung.com/br/.../centros-de-servico

Publications (papers)

Under Review **Unsupervised Quality Control of Natural Cork Stoppers Using Vision Transformer Autoencoders and Clustering**
Journal (Q1): Neural Computing and Applications
Authors: Marcos de Souza Oliveira, Jorge C. S. Cardoso, Vanessa Fernandes Ferreira, Armin Gharibi, André Perrotta, Licinio Roque, Antonio Silva Ferreira
Summary: Proposes a multiview transformer-based autoencoder for unsupervised cork stopper inspection. The model captures rich representations with low latency, outperforming classical methods in clustering accuracy and interpretability under industrial constraints.

In Progress **Addressing 5G Data Scarcity for Machine Learning Indoor Positioning via Digital Twins**
Venue: To be defined
Authors: Marcos de Souza Oliveira
Summary: This work proposes a 5G-based digital twin framework that integrates convolutional neural networks for real-time asset localization in industrial indoor environments. The approach achieves sub-meter accuracy and low latency, while also enabling the scalable generation of synthetic datasets to support future real-time location systems (RTLS) research and development.
URL: https://doi.org/10.6084/m9.figshare.30383902

In Progress **Graph-Based Interactive Visualization Tool for High-Dimensional Data**
Venue: To be defined
Authors: Marcos de Souza Oliveira
Summary: This work introduces an unsupervised visualization tool designed to explore high-dimensional data without prior labeling. Unlike traditional dimensionality reduction methods, it preserves key relationships in the original data. The tool offers two visualization modes: 1) graph-based and 2) similarity matrix, allowing interactive exploration of patterns and clusters. It enhances interpretability and supports the discovery of meaningful structures in complex data.
URL: https://doi.org/10.6084/m9.figshare.30383821.v1

2025 **A Study of Computational Vision Methods for Corrosion Detection in Industrial Assets**
Venue: 38th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)
Sub-venue: Workshop of Industry Applications (WIA)
Authors: Anderson G. Marco, Marcos de Souza Oliveira, Rodrigo Leite Prates, Murilo Cruz Lopes, Rogério Guedes Casal, Cristiani Vilela Ribeiro Guimarães
Summary: Comparative study of U-Net-based segmentation models for corrosion detection in industrial assets. InceptionV4 achieved the highest F1-score, while MobileNetV2 offered similar performance with reduced training time, highlighting its potential for real-time inspection systems.
DOI: In Progress

2025 **Deep Contrastive Variational Subspace Clustering**
Journal (Q1): Neurocomputing
Authors: Marcos de Souza Oliveira, Sergio Ricardo de Melo Queiroz, Cleber Zanchettin, Francisco de Assis Tenório de Carvalho
Summary: Introduces DCSSC, a deep clustering method combining contrastive learning and variational autoencoders to improve subspace clustering in high-dimensional data. The approach enhances relational structure and performs well with limited samples across 19 real-world datasets.
DOI: https://doi.org/10.1016/j.neucom.2025.130901

2024 **FastAiAlloc: A Real-Time Multi-Resources Allocation Framework Proposal Based on Predictive Model and Multiple Optimization Strategies**
Journal (Q1): Future Generation Computer Systems
Authors: Marcos de S Oliveira, Francisco Erivaldo Fernandes Junior, Lukas Cerveny, Flávia Akemi Miyazaki, Leonardo Valeriano Neri, Alan da Silva, Beatriz Leandro Bonafini, Victor Medeiros Outtes Alves, Órion Darshan Winter de Lima
Summary: Proposes a framework for real-time resource allocation in cloud platforms using multiple optimization approaches (GA, PSO, linear programming, and heuristics). The model also integrates user data in self-organizing maps to prevent over-allocation, achieving up to 60% performance improvement over baseline methods.
DOI: https://doi.org/10.1016/j.future.2023.08.014

2022 **Unsupervised Feature Selection Method Based on Iterative Similarity Graph Factorization and Clustering by Modularity**
Journal (Q1): Expert Systems with Applications
Authors: Marcos de S Oliveira, Sergio R de M Queiroz, Francisco de AT de Carvalho
Summary: Introduces KNMFS, a three-step unsupervised feature selection method using graph factorization and graph communities detection by Louvain heuristic. Demonstrates superior performance and efficiency across synthetic and real-world datasets.
DOI: https://doi.org/10.1016/j.eswa.2022.118092

2020 **Unsupervised Feature Selection Methodology for Clustering in High Dimensionality Datasets**
Journal: Revista de Informática Teórica e Aplicada
Authors: Marcos de Souza Oliveira, Sergio R de M Queiroz
Summary: Proposes a methodology to evaluate and combine unsupervised feature selection methods, addressing key questions on parameter tuning, feature ranking, and fusion strategies. Results show improved clustering performance in high-dimensional datasets.
DOI: https://doi.org/10.22456/2175-2745.96081

2020 **Predictive Model to Students Classification with University Retention**
Journal: Revista dos Mestrados Profissionais
Authors: Othon Luiz Teixeira de Oliveira, Marcos de Souza Oliveira, Orlando Bandrao, Eric Sales
Summary: Presents a predictive model for student dropout using CRISP-DM methodology and algorithms like logistic regression and decision trees. Achieved AUC of 0.80 and MAX_KS of 0.66 on a dataset of over 180,000 records.
DOI: https://doi.org/10.52614/2317-0115.2016.244962

Certifications

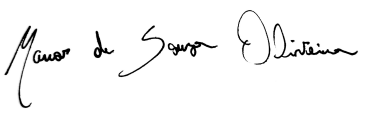
2024 Machine Learning in Production – DeepLearning.AI (Jun)
2024 Machine Learning Explainability – Kaggle (Jan)
2023 Business Analytics for Decision Making – Univ. of Colorado Boulder (Sep)
2023 Deep Learning for Business – Yonsei University (Aug)
2019 Mathematics for Machine Learning – Imperial College London
2019 Structuring Machine Learning Projects – DeepLearning.AI

Technologies

Python, pytorch, Keras, OpenCV, NetworkX, PySpark, Docker, Java, C++, SQL

Languages

Portuguese Native
English B2 (Upper Intermediate)



Coimbra, October 31, 2025 *Marcos de Souza Oliveira*