

MARCOS DE SOUZA OLIVEIRA

Senior AI Specialist | PhD Candidate | MSc

@marcosd3souza@gmail.com

<https://www.linkedin.com/in/marcos-de-souza-msc-893758aa>

Recife, Brazil



SUMMARY

I am an skilled IT professional with 10+ years working in the design and development of software and data-based solutions, providing to organizations high-value systems and platforms involving AI and data analytics with predictive and descriptive models. I've worked in different positions and agile projects with various team sizes, since distributed teams with 30+ people and also international projects.

My main interest is in discovery, design, planning and development of solutions that involve, or may involve, Artificial Intelligence and Machine Learning models and platforms (AI/ML Specialist).

I've strong academic background with achievements from PhD, MSc and Professor positions. I have also studied, experimented, developed and published papers in international high impact journals, focused on unsupervised machine learning and exploratory data analysis.

- * 5y+ AI/ML/DS positions
- * 5y+ Software Engineer
- * 1y+ ML instructor
- * 4+ AI/ML scientific publications

EXPERIENCE

Senior AI/ML Specialist

Instituto Nacional de Telecomunicações

2023 - Present Minas Gerais, Brazil

- Develop statistical and machine learning models using modern data science and machine learning techniques
- Develop applications for platforms that may involve AI/ML
- Design innovative solutions, using product discovery methodology, for Network as a Service (NaaS) platforms, with technologies such as Edge computing, IoT and AI/ML algorithms
- Conduct studies and analysis of AI/ML frameworks, data, platforms and tools
- Prepare technical documentation on proposed/adopted solutions
- Review developed code according to clean code and clean architecture practices

Data Scientist

CESAR - Innovation Institute

2019 - 2023 Recife, Brazil

The CESAR is an innovation center with more than 27 years of expertise and multidisciplinary projects on different technological fronts, such as IoT, AI and software engineering. As a Data Scientist my main responsibilities were

- Modeling data-based solutions to meet requirements in terms of scalability, low cost and security for medium, large and international projects
- Through an internal initiative involving LLM models, I experimented and developed a tool capable of automatically generating technical reports for use in projects
- Within a data-driven squad, I proposed a probabilistic model for predicting employee turnover. The idea is to provide decision support to the organization C-level in order to understand motivations and propose initiatives to control voluntary turnover
- Development of backlog and planning for short, medium and long term in order to attend innovative features, as well as in the writing of technical reports that justify government incentives for the development of research and development (R&D) projects
- Development of anomaly detection models to prevent and identify outliers/crashes in applications that observe data consumption over time (time series)
- Development of models and performance metrics (KPIs) for allocating resources on the platform in a real-time scenario. This work was presented to Dell and the success resulted in development of a published paper in an important international journal
- As part of the Samsung customer platform, I have developed data-driven solutions for Samsung assistance. The main results achieved were a route suggestion system for technicians, in order to decrease traveled distance and to meet all the service demand and also the development of optimization models to perform the best organization of queues in Samsung's assistance. Studies on queuing theory and the Erlang method were also considered

EXPERIENCE

Machine Learning Instructor

CESAR School

📅 2022 - 2023 📍 Recife, Brazil

CESAR School is the teaching branch of the CESAR innovation institute. The variety of courses and formats, such as bootcamp, degree, master degree and also professional doctorate are offered in order to solve real problems of clients and partners, using a problem-based learning approach (PBL)

- As a facilitator, I participate in several initiatives such as the accelerated training course in TechDesign (FAST), which aims to reinsert professionals in key areas such as design, testing and machine learning with the most varied backgrounds (whether they are technology-based or not)
- As an instructor in the Data Specialization, I teach the Unsupervised Learning discipline. Where students will have an immersion on the main topics of the unsupervised area, such as the approach of unsupervised techniques to dimensionality reduction, classic clustering algorithms, unsupervised metrics to performance evaluation and the introduction to deep clustering
- I help the undergraduate students conducting software-based projects and solutions that apply machine learning and design fundamentals to solve real problems demands by clients/partners

Software Engineer

CESAR - Innovation Institute

📅 2017 - 2018 📍 Recife, Brazil

- Design and development of ETL process attending low cost requirements to average waiting time estimation API using AWS components
- Design and implementation of the queue management system simulator (timeline queue simulator) to perform data ingestion and pre-processing for ML/AI models
- Demonstrated multidisciplinary at software development to consumers queue manager, working on frontend (AngularJS) and backend (NodeJS) features with docker containers on AWS EC2
- Development of android engine (AAR) to noise reduction (low/high/band pass audio filter) on voice audio

Android Developer

Stefanini

📅 2016 - 2017 📍 Recife, Brazil

Document Solutions (DS) was a Stefanini affiliate focused on capturing, processing and recognizing documents such as bank checks and the Brazilian identity register. From the document solution, two new businesses emerged: di2win and Stefanini dn.ia. In my DS journey, the following results was achieved

- Development of an application for opening digital bank accounts from scratch. The success of this PoC led a important Brazilian bank (Itaú) to request that we take over the mobile apps area. However, due to the size of the team, we didn't take over this department
- Integration into an android application by JNI and improvement of image segmentation engines (applied research) from code written in C++
- Provision of engines, in AAR format, for use in applications developed by BBVA Argentina. We had meetings directly with those responsible for the bank's technological area to provide details and demands
- Portability of the bank check capture and recognition engine (DSPhoto) on a variety of smartphones, from widely recognized manufacturers such as Samsung and Apple, to manufacturers not so well known at the time such as Xiaomi

Software Engineer

Samsung Electronics (UFPE)

📅 2015 - 2016 📍 Recife, Brazil

The Samsung/Cin-UFPE Project is a research branch of Samsung Electronics within the Federal University of Pernambuco (UFPE), which aims to develop R&D solutions to be ported, delivered and consumed by Samsung end users around the world. For this reason, the projects developed under the Samung/Cin-UFPE Project must meet the code pattern, requirements, software quality and documentation required for any Samsung solution. The project is divided into two main areas: MAC and ITS. MAC refers to solutions for mobile or wearable products, such as smartphones, and ITS focuses on solutions for TVs and notebooks in general. My main activities within MAC were

- I participated in the development of an API that would serve as a backend for a Samsung social network projected to be made available at the Olympic games in Rio 2016. The idea was for this API to support users and events registration. We had consultancy from UFPE to meet security, scalability and integration requirements with other social networks using OAuth
- I also took part in the research and development of an application aimed at the visually impaired. The idea was to develop and make available a guideline for any application wishing to meet accessibility requirements, such as the size and positioning of buttons and programming more assertive messages from the Google Talkback assistant
- I took part in an initiative aimed at developing cross-project tools. The idea was to develop utility applications that checked that the requirements made by Samsung were met in the projects, such as checking that all the code produced was written in English, the presence of header copyright in all possible files and performance tests for high availability requirements

EXPERIENCE

Java Developer

MV

📅 2014 - 2015 📍 Recife, Brazil

MV is a health tech company that deals with all fronts for the operation of large hospitals throughout Brazil, whether public or private. In addition to the various products offered by MV, two stand out. The first is based on Oracle forms technology (MV2000) and the second is based on Java web (SoulMV). Both have all the modules to ensure hospital operations. I was assigned to the supply module team. As a software developer, I have been involved in

- Maintaining legacy systems
- Migrating technology platforms (e.g migrate PL/SQL Oracle Forms to Soulmv Java Web)
- Developing new functionalities using the Java web

EDUCATION

PhD: Machine Learning

Federal University of Pernambuco, Brazil

📅 2020 - 2024

- Deep Subspace Clustering Approach Based on Generative Learning and Tensor/Matrix Decomposition

GPA
3.6 / 4.0

MSc: Machine Learning

Federal University of Pernambuco, Brazil

📅 2016 - 2018

- In this work we propose a methodology that to aid data specialists to answer important questions: (1) do current feature selection methods give similar results?(2) is there is a consistently better method?(3) how to select the m-best features?(4) as the methods are not parameter-free, how to choose the best parameters in the unsupervised scenario? and (5) given different options of selection, could we get better results if we fusion the results of the methods? If yes, how can we combine the results? We analyze these issues and propose a methodology that, based on some unsupervised methods, will make feature selection using strategies that turn the execution of the process fully automatic and unsupervised, in high-dimensional datasets

Specialization: Software Engineering

University of Pernambuco, Brazil

📅 2015 - 2016

- This specialization focuses on developing the main fundamentals and good practices in software development. Addressing important disciplines such as requirements engineering, interface and user experience (UX/UI), software project management and architecture, agility and quality assurance (QA) process

BSc: Computer Science

Faculdade dos Guararapes

📅 2011 - 2014

GPA
5.16 / 6.0

KEY ACCOMPLISHMENTS

Data-driven business mindset

that helped to understand and propose data-driven solutions with high business value

Publication of paper in high-impact international journal

that was developed inside Dell project and implemented in Dell AI Platform.

Optimized & Refactored queue management system

that resulted in meet a high consumer demand for Samsung technical assistance