Marcos de Souza Oliveira

📍 Recife, Brazil   ✉ marcosd3souza@gmail.com   📞 +55 081 98784-4363   🔗 marcosd3souza.github.io

in marcos-d-souza   ○ marcosd3souza

Summary

I am a highly skilled researcher and professional with over 13 years of experience in software development and Artificial Intelligence (AI), Machine Learning (ML), and data analytics. I have led and contributed to a wide range of projects, from agile development teams to large-scale international collaborations, holding roles such as Tech Lead, AI/ML specialist, Data Scientist, and Software Engineer. I am particularly drawn to unsupervised learning and exploratory data analysis.

Interests

- Exploratory data analysis
- Cluster analysis
- Deep clustering

Education

PhD	**Federal University of Pernambuco - UFPE**, Machine Learning **Area:** Deep clustering	2020 – 2024
MSc	**Federal University of Pernambuco - UFPE**, Machine Learning **Area:** Unsupervised Feature Selection	2016 – 2018
Spec	**University of Pernambuco**, Software Engineering	2015 – 2016
BS	**Faculdade dos Guararapes**, Computer Science	2011 – 2014

Relevant Experience

Inatel, AI/ML Specialist — Minas Gerais, Brazil (remote) — 2023 - current

- Development of statistical and AI/ML models applied to industrial process
- Prospect new AI/ML projects by writing pre-proposal/pre-sale and work plan documents
- Conduct studies and analysis of AI/ML frameworks, data, platforms and tools
- Publish results in top conference/journals (Scopus indexing)

CESAR School, Machine Learning Instructor — Recife, Pernambuco, Brazil — 2022 - 2023

- dimensionality reduction
- classic clustering algorithms,
- unsupervised metrics
- deep clustering.

CESAR, Data Scientist — Recife, Pernambuco, Brazil — 2019 - 2023

- Development of explainable classifier (tree-based) for predicting employee turnover
- Development of anomaly detection models to identify outliers/crashes in applications that observe data consumption over time (time series)
- Design and implementation of the queue system simulator to perform data ingestion for AI/ML models
- Development of route suggestion system based on clustering for Samsung technicians (optimization of traveled distance to meet all demanded attendances)
- Development of optimization models to perform the best organization of queues

in Samsung's assistance by evolutionary algorithms (genetic / PSO)

CESAR, Software Engineer   Recife, Pernambuco, Brazil
                                                                         2018
- Design and implementation of the queue system simulator to perform data ingestion for AI/ML models
- Implementation of NLP classifier to detect furious feedback in the consumers text messages (Samsung concierge)

Stefanini, Software Engineer   Recife, Pernambuco, Brazil
                                                                         2017
- Developing mobile applications for pattern recognition engines (image document segmentation).

Informatics Center, UFPE, Machine Learning Scholarship   Recife, Pernambuco, Brazil
                                                                         2016
- Dedicated to research work for the master's degree in Machine Learning
- Unsupervised Feature Selection Framework

Publications

Deep Contrastive Variational Subspace Clustering   2024

Oliveira, M. D. S., Queiroz, S. R. M., Zanchettin, Cleber., Carvalho, F. A. T.

(Under Review)

FastAiAlloc: A real-time multi-resources allocation framework proposal based on predictive model and multiple optimization strategies   2023

Oliveira, M. D. S., et al.

Future Generation Computer Systems

10.1016/j.future.2023.08.014 ↗

Unsupervised feature selection method based on iterative similarity graph factorization and clustering by modularity   2022

Oliveira, M. D. S., Queiroz, S. R. M., Carvalho, F. A. T.

Expert Systems with Applications

10.1016/j.eswa.2022.118092 ↗

Unsupervised feature selection methodology for clustering in high dimensionality datasets   2020

Oliveira, M. D. S., Queiroz, S. R. M.

Revista de Informática Teórica e Aplicada

10.22456/2175-2745.96081 ↗

Projects

Cluster Viz Tool   dataviz repo ↗
- Developed a dataviz tool for cluster analysis of high-dimensional datasets using neighborhood graph and similarity matrix
- Technologies Used: JavaScript, D3

Technologies

Languages: Python, Matlab, Java, SQL, JavaScript

Technologies: Scikit-learn, Pytorch, Tensorflow, NetworkX