

Daniel Pereira da Costa

danielp3@usc.edu | linkedin.com/in/daniel-pereira-da-costa | danieldacosta.github.io | github.com/DanielDaCosta
Los Angeles, California | (213) 732-6265

EDUCATION

University of Southern California

Master of Science in Applied Data Science

Los Angeles, United States

August 2022-May 2024

École Centrale Paris

Master of Engineering - Double Degree with PUC-Rio

Paris, France

September 2016-December 2019

Pontifical Catholic University of Rio de Janeiro (PUC-Rio)

Bachelor in Control and Automation Engineering, Minor in Mathematics

Rio de Janeiro, Brazil

January 2014-December 2019

TECHNICAL SKILLS

Programming Python (Pandas, Numpy, PySpark, Scikit-learn), SQL (PostgreSQL, MySQL, SQL Server), MongoDB, JavaScript, C/C++

Cloud Platform Amazon Web Services (AWS)

ETL Tools Airflow, AWS Step Functions, AWS Glue, Amazon Elastic MapReduce

Other Tools Power BI, Tableau, Apache Spark, Pytorch, TensorFlow, Terraform, Gitlab CI/CD, Docker, Git

WORK EXPERIENCE

Intel Corporation

May 2023-Present

Data Analyst

Folsom, United States

- Develop an all-encompassing PowerBI dashboard for proactively identifying performance deviations in 50+ AI workloads running on Intel's Ponte Vecchio GPU across single-card, multi-card, and multi-node configurations.
- Design a PoC using GPT-4 and Llama-Index to ingest and associate unstructured customer engagement updates from disparate sources (e.g., PowerPoint, JIRA, Excel) into a single PowerBI based reporting for several high priority, high visibility engagements.

January 2022-August 2022

Akad Seguros

Data Engineer

Rio de Janeiro, Brazil

- Orchestrated design of 600GB Data Lake on AWS through S3, Glue, Lambda, and Step Functions; published to [AWS Blog Brazil](#).
- Ingested over 100GB of data from 8 different on-premises and cloud-based servers into Data Lake via Data Migration Service and ETL jobs.
- Defined and deployed Data Warehouse (50GB) components and architecture on AWS applying tools such as S3 for storage, Spark for processing, and Athena for analytics.

Cyberlabs (Google for Startups 2020)

January 2020-December 2021

Data Scientist

Rio de Janeiro, Brazil

- Engineered Machine Learning model with 71% f1-score to reduce churn from 6 million user cybersecurity app; deployed model in AWS utilizing EMR Clusters with Spark for data preprocessing, EC2 instance for ML inference, and Data Pipeline for ETL.
- Guided a team of 4 in creating a Churn Predictive Model (LSTM network) with 72% precision and 83% recall for telecom company; leveraged statistical analytics techniques to analyze customer churn patterns against 10 key metrics.
- Spearheaded a team of 3 in constructing a system aiding 20+ companies in controlling COVID-19 spread for 3,000 employees; implemented event-driven architecture on AWS employing Infrastructure as Code and Lambda, SQS, SNS, API Gateway, and RDS.
- Collaborated with 2 members to construct an XGBoost model with 77% f1-score to maximize customers' revenue.

Cyberlabs (Google for Startups 2020)

July 2019-December 2019

Data Scientist

Rio de Janeiro, Brazil

- Built Node.js back-end service generating up to 10 insights per minute from Computer Vision Model; increased 100 establishments' decision-making productivity by 20 hours/month.
- Integrated and deployed microservice with GraphQL and Docker.

Softbank Robotics Europe

June 2018-December 2018

Embedded Software Engineer

Paris, France

- Created firmware module in C/C++ embedded in a microcontroller for a robot's inductive sensor (LDC1312/4).
- Presented reports to firmware team to validate sensor's effectiveness against 2 KPIs.

PROJECTS

Evaluation of Nonlinear System Identification to Model Piezoacoustic Transmission

[IFAC 21st World Congress](#)

Artificial Neural Networks, Nonlinear System Identification, Mathematical Modeling, and Smart Structures

- Developed data-driven black-box models (ARMAX, NARX, and Artificial Neural Network) for an acoustic transmission modeling problem; model hyperparameters tuned by Grid Search, achieving R2 Free-run Simulation (FRS) of 92%.

Machine Learning pipeline with Airflow and Elastic Container Service

[Towards Data Science](#)

Amazon Web Services, Airflow, Docker, Elastic Container Service (Fargate), Elastic Container Registry

- Created an Airflow pipeline and deployed containerized application to a serverless compute engine (Fargate) for machine learning inference; published tutorial on Towards Data Science.