

LUIZ FERNANDO COSTA DOS SANTOS

luizfernando1012000@gmail.com

linkedin.com/in/luiz-fernando632

github.com/luiz826

luiz826.github.io

Languages Portuguese, English

Experience

Nubank

Machine Learning Engineer - ML Platform team

Sep 2025 – Present

São Paulo, Brazil (Remote)

- Spearheaded the development of the Feature Store, standardizing feature creation, discovery, and versioning, which reduced the time-to-market for new models.
- Developed new feature API endpoints (for discovery by country and namespace), optimizing the development workflow and reducing integration time.
- Architected data immutability systems (e.g., using Delta Lake, versioning) to enforce data quality standards and ensure the reproducibility of feature pipelines.

Nubank

Machine Learning Engineer - Credit team

Sep 2022 – Present

São Paulo, Brazil (Hybrid)

- Led Python model orchestration (batch/real-time) library, ensuring seamless integration and reducing model deployment time from 5 months in avg to 2 months.
- Developed and deployed a flexible binning framework for model score management across multiple versions, reducing binning change cycles from approximately 3 months to a few weeks.
- Led the design and migration of monitoring dataset services to a new Spark Scala framework, creating robust abstractions for each monitoring type, reducing contribution times from +5 hours to less than one hour.
- Migrated ETL to Data Mesh, cutting infrastructure costs by 25% and improving governance. The team had almost 5K datasets migrated.

Huawei & Embrapii

Machine Learning Engineer

Feb 2022 – Sep 2022

João Pessoa, Brazil

- Embrapii is a company that allow Brazilian Universities easily work together with large companies, in this case, Huawei.
- Built a feature serving system, with Airflow, gathering data from IoT and sensor devices, transforming and storing data in a Influx DB, with a latency of less than 500 ms.
- Developed a solar characterization dataset for optimal solar tracker angles with more than 50K rows from several interested regions.
- Trained a Random Forest model, with Scikit Learn, improving energy generation by up to 12%.

Education

B.S. in Data Science and Artificial Intelligence - Universidade Federal da Paraíba

Aug 2020 – Jun 2025

- GPA: 9.43/10.0 (2nd of the class)
- TF for Data Science Introduction, with a +30 students class

M.S. in Computer Science Candidate - Georgia Tech

Aug 2026 – Aug 2028

- Applying for the OMSCS in the next available semester.

Technical Skills

Languages:

- Scala, Python, SQL, Clojure, R, Java, C, C++

Frameworks:

- Spark, Databricks, AWS (S3, EMR, Glue, EC2), Datomic, PostgreSQL, InfluxDB, Kubernetes, Kubeflow, Docker, Airflow, Terraform, CI/CD, Tekton, Grafana, PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, XGBoost

Certifications

Machine Learning in Production – DeepLearning.AI

2025

- Course taught by Andrew Ng, from Stanford

Neural Networks and Deep Learning – DeepLearning.AI

2021

- Course taught by Andrew Ng, from Stanford

CS50 – Harvard

2025

- As the final project, developed a Computer Vision model to detect a rubber duck in a image using Yolo, and built a front-end usign HTML, CSS and Java Script. The model had 97% of AUC.