Daniel Lima

Senior Data Scientist | Machine Learning

Personal Info

Email
daniel.vtlima@gmail.com

LinkedIn
www.linkedin.com/in/danielvictor94/

Phone
+55 91 98841-3504

Skills

Machine Learning (XGBoost , LightGBM , SKLearn)

Python , Pyspark , SQL

Time Series , ARIMA , SARIMAX , Prophet AutoML

Recommender Systems , LightFM

Statistical Data Analysis

Genetic Algorithms

Containers - Docker , Kubernetes

Machine Learning Deployment

Azure Databricks

AWS Services

Snowflake

Airflow

Data Transformations Pipelines

ETL / ELT

Deep Learning (CNN)

Data Visualization - Tableau , Dash , Streamlit

Object Oriented Programming (OOP)

Survival Analysis

Machine Learning Solutions

Linux

Agile SCRUM

Communication

Technical Lead

A Data Professional (8+ years) with a Master's in Electrical Engineering with a focus on Machine Learning and a background in Computer Engineering Bachelor. Experience in leading technical teams into building end-to-end Machine Learning Solutions and translating stakeholders' needs into data products.

Work History

2024-05 - present

AI Engineer | Data Engineer

834 Labs, Remote

- Build ETL pipelines using AWS Glue in conjunction with Managed Airflow to process and orchestrate healthcare data according to business needs;
- Queries optimization and Data analysis using SQL;
- Data modeling and data retrieval automation;
- Collaborated with system architects, design analysts, and others to understand business and industry requirements.
| Stack: Data Pipeline, AWS (Glue, S3, VPC, Redshift, Managed Airflow), Data Transformation, Data Quality, Data Lake, Data Warehouse, Data Modeling, Data Base Management (CREATE, and UPDATE tables), APIs, Python, Automation, SQL, Data Orchestration.

2023-06 - 2024-06

Staff Data Scientist

TOTVS Labs, Remote

- Utilized advanced querying, visualization, and analytics tools to analyze and process complex data sets, translating business needs into long-term data models.
- Applied statistical and algebraic techniques to interpret key insights from collected data, with a focus on building efficient solutions.
- Improved model performance by applying various code techniques, ensuring scalability and long-term efficiency, reducing the time and resources by more than 50%.
- Developed and implemented mathematical techniques to enhance engineering and scientific solutions.
- Set up SQL database on cloud services (GCP) to store client data for further data analysis.
- Leveraged time series analysis to identify trends make seasonality adjustments, and forecast ROI in advertising campaigns.
| Stack: Code Performance and Quality, GCP (Big query), XGBoost, Lightgbm, Data and Client Discovery, Big data analysis, Data Modeling, Data Base Management (CREATE, and UPDATE tables), Python, SQL, Metaheuristics, OOP, Git, Apache Beam.

2023-07 - 2024-01

Senior Data Consultant

Customer Times, Remote

- Collaborated on ETL (Extract, Transform, Load) tasks, maintaining data integrity and verifying pipeline stability.
- Prepared documentation and analytic reports, delivering summarized results, analysis, and conclusions to stakeholders.
- Monitored incoming data analytics requests and distributed results to support healthcare strategies.
- Analyzed complex data and identified anomalies, trends, and risks to provide useful insights to improve internal controls.
- Built databases and table structures for Machine Learning and analytics applications.
- Collaborated with system architects, design analysts, and others to understand business and industry requirements.
| Stack: Data Pipeline, AWS (Glue, Lambda, S3, Redshift), Data

Languages

English ★★★★★

Portuguese ★★★★★

Master's Dissertation and Paper

- Heuristic for the Optimization of Data Communication Networks as Support for the Implementation of Smart Grids

- Optimization of Data Communication Networks as Support for Smart Grid Deployment.

Transformation, Data quality, Data Lake, Data Warehouse, Data Modeling, Data Base Management (CREATE, and UPDATE tables), APIs, Python, Spark, SQL, Airflow, Data Orchestration, OOP, Unit tests, Integration Tests, Postman

2022-12 - 2023-07

Senior Data Scientist

Ioasys | Alpagartas, Remote

- Developed advanced machine learning models and interactive dashboards using Tableau, Prophet AutoML, LightGBM, HyperOPT, Fugue Tune, and ARIMA to effectively communicate business insights gained from time series models;
- Advocated for software engineering best practices and utilized OOP, Pyspark, and Databricks to ensure efficient and effective analysis of large datasets;
- Refactored legacy code using OOP principles and Python, Spark, and Databricks to deliver high-quality, efficient, and maintainable code for production readiness;
- Created and maintained up-to-date and accurate dashboards to enable data-driven decision-making, collaborating closely with stakeholders throughout the project lifecycle.

| **Stack**: Databricks , AWS , Snowflake, Prophet AutoML, Python, Pyspark, SQL, ARIMA

2021-10 - 2022-12

Senior Data Scientist

Farmbits, Remote

- Developed cutting-edge AI and ML applications: designed and built custom dashboards and Machine Learning (LightFM) hybrid recommendation systems and integrated them into the AWS cloud infrastructure through API (FlaskAPI), following software development best practices;
- Held regular meetings with clients to understand their challenges and recommend comprehensive data-driven solutions;
- Ensured that all applications were deployed efficiently into production, meeting all performance and quality standards;
- Delivered actionable insights and data-driven solutions that improved business operations and increased revenue by 20% as a quick win;
- Be the technical reference for my Data team and be the AI reference for the company.

| **Stack** : LightFM , LightGBM, Dash, Python, AWS Services (EC2 , S3, EKS , AuroraDB , Sagemaker), Postgres, Kubernetes, Docker , Jenkins (CI/CD)

2021-04 - 2022-04

AI Resident | Data Scientist and Machine Learning Engineer

Fiep, Remote

- Held regular meetings with clients to understand their challenges and recommend comprehensive data-driven solutions;
- Led the development of two projects, including clustering failure patterns of a machine in a high production environment and ranking components to improve maintenance schedule efficiency;
- Predicted the amount of moving capital in a given period to enhance cash flow planning. Utilized the SARIMA model to estimate the values of the moving capital time series in the second sprint, where I built a recommender system to improve an online store's sales and competitiveness;
- Developed a recommendation system to enhance the sales of an Agro company and created a Genetic Algorithm to optimize the production of a home appliance line;
- Concluded the program by investigating contactless fingerprint field computer vision-related methods to determine the best course of action to

begin the development phase.
| Stack : Convolutional Neural Networks (CNN), Genetic Algorithms, XGBoost, Python, PyTorch, LightFM, SARIMAX, Streamlit, Flask API, Docker.

2019-10 - 2021-04   Data Scientist | Data Engineer

Freelance, Remote

- Automated ETL processes using Python and Airflow to transform and load data into the data lake (S3 Buckets)
- Created interactive visualizations and dashboards in Tableau and Dash to help stakeholders gain insights from the data.
- Containerized the data visualization application using Docker and deployed it across different environments (DEV, PRD)
- Built data reports with data analysis and data visualization to provide valuable insights into business performance and customer behavior.
- Implemented an ELT process to analyze customer feedback data from multiple sources.
- Database Management with Postgres
| Stack: Database Management, Data Lake Management , Data Transformations Pipelines , Machine Learning (Sentiment Analysis) , NLP , Data Analysis , Python , Data Viz , Tableau , Dash , Docker , ETL / ELT , Airflow , FlaskAPI , BI report , Databricks

2019-04 - 2019-10   Full-stack Lead Developer

Equilibrium_Web, Belém-PA

Responsible for managing systems and developing new software solutions to meet the needs of the organization. My role involved using my expertise in software development to create efficient and effective solutions that improved the functioning of the organization. Through my efforts, I was able to make a significant impact in improving the overall performance and productivity of the public defense organ.
Main responsibilities :
 - Be the Lead of a team of three developers and was responsible for creating end-to-end solutions to meet stakeholder requirements. My responsibilities included translating technical results to business owners, and ensuring that their needs were met in a timely and efficient manner.
- Leveraged my skills in identifying business pain points to create effective technical solutions that addressed the underlying issues. Through my leadership and technical expertise, I was able to deliver high-quality solutions that added significant value to the organization.
- Do the Database Management, by creating and maintaining the public organ data (create and update tables using SQL and OracleDB)
| Stack: Team Management, Leadership, C# , SQL , Oracle DB , Database Management , OOP

2017-09 - 2019-04   Research and Development Data Scientist (R&D)

Universidade Federal do Pará (UFPA), Castanhal-PA

As a data scientist, I was responsible for conducting the Research and Development of image processing and computer vision techniques to evaluate and classify ovine cuts based on their conformity and quality levels. My responsibilities involved developing and implementing pre-processing phases related to Edge Detection methods and analyzing conformity data to feed the Machine Learning model. Through my expertise in data analysis and computer vision, I was able to create a robust solution that delivered accurate results for ovine cuts classification.
| Stack: Machine Learning, Computer Vision, Image Processing, Edge Detection, Matlab.

2015-06 - 2017-12	**Research Data Scientist** *Universidade Federal do Pará (UFPA), Castanhal-PA* As a Research Data Scientist, I developed a Kinect application aimed at improving the rehabilitation process for stroke patients. Through the use of Biofeedback, the application helped enhance the acceptance of treatment by the diseased by using gamification of the treatment. By leveraging my skills in data analysis and software development, I was able to create an effective solution that had a positive impact on patient outcomes. | **Stack**: Healthcare data , Biofeedback , C# , Kinect , Computer Vision , , gamification.
2015-09 - 2016-09	**Research Data Scientist Assistant** *Universidade Federal do Pará (UFPA), Castanhal-PA* As a Research Data Scientist, I developed an automated software testing algorithm for the WEB-PerformChart framework. This involved utilizing C++ and MySQL to build a Markov Chain model that represented each state of the program. By doing so, I was able to identify potential errors caused by conflicts between different states of the software, thereby ensuring its optimal performance. | **Stack**: Markov Chains , C++ , HTML , CSS.

Education

2021-04 - 2022-04	**AI applied to Industry, Specialization** *Fiep, : Convolutional Neural Networks (CNN), Genetic Algorithms, Time Series Analysis, Deep Learning*
2018-09 - 2020-12	**Applied Machine Learning, Master's Degree** *Universidade Federal Pará, : Graph Theory, Linear Algebra, Machine Learning, Classic Optimization Methods, Metaheuristics*
2013-04 - 2020-01	**Computer Engineering, Bachelor's Degree** *Universidade Federal do Pará, : Computer Vision, Machine Learning, Markov Chains*

Courses

2022-07	**AWS Sagemaker: Machine Learning in Production.**
2022-06	**Apache Beam: Data Pipeline with Python**
2022-05	**MLOPS: Machine Learning and API's**
2022-05	**Vagrant: Managing Virtual Machines.**
2022-02	**Kubernetes: Pods, Services and Configmaps**
2022-02	**Deploy in Amazon EC2: High Availability and Scalability of an Application.**