

WORK EXPERIENCE

Serasa Experian (credit bureau) Latin America DataLab	Sao Paulo, Brazil
Staff Data Scientist	09/2023 - Present
<ul style="list-style-type: none">For the past 4 months, I have led GenAI implementations across Serasa as the team's tech lead. Coordinating a team of four individual contributors in building NLP applications such as a legal copilot and a custom email campaign generator. Leveraged the use of retrieval augmented generation minimizing hallucination.	
Serasa Experian (credit bureau) Latin America DataLab	Sao Paulo, Brazil
Senior Data Scientist	05/2023 - 08/2023
<ul style="list-style-type: none">Introduced a MLOps template using Docker and Airflow for automated client backtesting.Developed an insurance claims forecasting framework that outperformed the previous model by over 10 ROCAUC points, surpassing industry benchmarks.	
Serasa Experian (credit bureau) Latin America DataLab	Sao Paulo, Brazil
Data Scientist II	11/2021 - 04/2022
<ul style="list-style-type: none">Led a team of two junior data scientists in developing a credit score for the agricultural sector resulting in a product that raises 5M USD annually. Defined target population and designed business-centric metrics. Replaced external data source, improving model autonomy without sacrificing performance.	
Itaú Unibanco (bank) Advanced Analytical Consulting Management	Sao Paulo, Brazil
Data Scientist I	11/2020 - 10/2021
<ul style="list-style-type: none">Led a team of two data science interns in developing a model to set monthly sales targets to Itaú's bank officers using causal inference uplift/heterogeneous treatment effect models. A/B testing validated projected 3% annual revenue increase for tested products.	
Itaú Unibanco (bank) Advanced Analytical Consulting Management	Sao Paulo, Brazil
Data Science Intern	10/2019 - 10/2020
<ul style="list-style-type: none">Designed causal inference model for pricing, assessing Itaú portfolio elasticity. Introduced a new pricing dimension, impacting 30M customers.Enhanced business processes using pandas, SQL and Hive, reducing operation time.	

EDUCATION

University of São Paulo (USP)	08/2021 - 07/2024
Master's degree in Probability and Statistics	GPA 4/4
Federal University of ABC (UFABC)	05/2015 - 04/2021
Bachelor's degree in Science and Technology (Pure Mathematics)	GPA 3.88/4

OPEN KNOWLEDGE AND OPEN SOURCE SOFTWARE

Machine Learning blog	vitaliset.github.io
Vitali Set is my personal site where I share insights and knowledge about Machine Learning.	
scikit-learn	All my pull requests to scikit-learn
datalib	All my pull requests to datalib

TOOLS I USE DAILY

functional and oop with python, numpy, matplotlib, scikit-learn, pytorch, xgboost, langchain, pandas, sql, pyspark, git, pytest, ci/cd, airflow, docker, linux, azure.

RESEARCH FELLOWSHIPS

Wada Property in Doubly Transient Chaos	11/2018 - 09/2019
FAPESP (São Paulo Research Foundation)	
The Euler-Maclaurin formula and a few applications in Pure and Applied Mathematics	09/2017 - 08/2018
FAPESP (São Paulo Research Foundation)	
Numerical study of orbits in the 3-body problem	08/2016 - 07/2017
CNPQ (Brazilian National Council for Research and Technological Development)	
Introduction to stellar structure and evolution	08/2015 - 07/2016
UFABC, Researching Since the First Day Program	