

Gabriel Tardochi Salles

São Paulo, São Paulo, Brazil

ga.tardochisalles@gmail.com

linkedin.com/in/gabrieltardochisalles

Summary

Data scientist with solid software engineering and cloud computing fundamentals. I have four years of experience developing machine learning (especially deep learning) solutions that deliver business value. My career is driven by a passion for solving complex problems, particularly in the domains of natural language processing/ understanding and computer vision.

>> gabrieltardochi.medium.com: I regularly share some (hopefully) interesting data science stuff on Medium.

Experience



Senior Data Scientist

Lett

Apr 2023 - Present (9 months)

DataOps

- Integration and utilization of Large Language Models (LLMs, such as GPT-4) in data annotation workflows via active learning, enabling the annotation of thousands of samples in a very short time, reducing the time to production by 3x for models that relied on new data annotations, and also decreasing annotation costs by 20x;
- Development of a named entity recognition model, achieving a token-level accuracy of 93%;
- Responsible for our MLOps stack and pipelines;
- Maintenance of almost ten machine learning solutions in production;
- Creation of a multitask model for natural language understanding and processing to enrich data, quadrupling our deduplicated product database;
- Applied survival analysis to estimate the time until product page change, avoiding inefficient data crawling - a 40% reduction in data capture costs without significant loss of data quality;
- Developed an anomaly detection system behind a new data transparency product for our clients, positively highlighting our data quality, a vital pillar of the company's strategy;
- Conducted a market study measuring the influence of positioning well in e-commerce searches on sales, which is highly relevant to large industries looking to increase their market shares.



Data Scientist

Lett

Nov 2021 - Apr 2023 (1 year 6 months)

DataOps

- Reduced operations product matching demand by 96%, increased efficiency by 5x through AI-based product filtering and matching with state-of-the-art Deep Learning approaches.
- Reduced operations product image evaluation demand by 50% by detecting negligible image changes with Computer Vision algorithms.
- Increased human-fault Data Quality assertiveness from 85% to 99%, establishing DQ as a notable Lett differential, by applying anomaly detection ML techniques.
- Created a scalable and cost-effective product images' quality evaluation system, bringing to life the most requested feature by our customers, at the time.

- Actively helped product teams and business decision makers understand opportunities that data science and machine learning could unlock.
- Built and implemented cloud infrastructure, tooling and frameworks to achieve faster model development and deployment, while ensuring MLOps principles in a Big Data scenario.



Junior Data Scientist

Lett

May 2021 - Nov 2021 (7 months)

Scale & Operations

- Scaled Lett's EQI (E-commerce Quality Index) study to become an industry benchmark, by applying state-of-the-art Natural Language Processing to automatically extract products information, such as brand and categories.
- Improved Operations product matching capabilities by 4x through a BERT-based semantic search engine.



Data Science Intern

DSM

Mar 2020 - May 2021 (1 year 3 months)

- Reduced last mile distribution costs by 15% by developing a custom IRP (Inventory Routing Problem) solver to optimize planning under both inventory and fleet unavailability.
- Assisted and taught colleagues as a "Digital Guide" on a global digitalization movement. Even acted in a funny video series about digital tools.
- Applied greedy algorithms to optimize distribution centers' inbound performance.
- Used machine learning to cluster customers based on their region, truck size acceptability, receiving restrictions and priority.
- Collected, studied, and interpreted large datasets to generate reports and dashboards.

Education



Universidade Presbiteriana Mackenzie

Bachelor's degree, Computer Science

Jul 2018 - Jul 2022

- Participated in Artificial Intelligence academic leagues.
- Acted as an Algorithms and programming monitor.
- Capstone project: Emotion Recognition in Argument Mining with Deep Learning.
- BSc Computer Science. GPA: 3,7.

Licenses & Certifications



Deep Neural Networks with PyTorch by IBM (with Honors) - Coursera

SUSMBDE3HZU6




Intermediate Python - DataCamp


471e3a06591a7bcaa29243d5a88e7cbce0ae6a28





Introduction to Data Visualization with Matplotlib - DataCamp

05e01b4566f874e3e4adcdb6300acd5d27f9cadf


 **Unsupervised Learning in Python** - DataCamp
bcbad3b3dddc28ef18bbce00399f536ae20dc59d


 **Data Science for Business** - DataCamp
af26a335f24d596030672f053cbb3d670d9796f2


 **Data Manipulation with pandas** - DataCamp
3b8e17de2bbbd024074311231073083d93293790


 **Supervised Learning with scikit-learn** - DataCamp
48b5754dba9f1fdca093c156ca0541a52d840d01


 **Joining Data with pandas** - DataCamp
d17a7ab04ff467813e897bc71bd9b5e91e44479b


 **How to Build and Deploy a Complete Data Science Solution** - Mario Filho - Data Science
721ac335-617e-4d05-8fd3-f3e052479359

 **Winning a Kaggle Competition in Python** - DataCamp
1210ffe55768196b5964b7d550651cef5c633341

 **Introduction to Spark SQL in Python** - DataCamp
f5d403b8493258495fdf5fb7211bed0b1b6bdd4c

 **Customer Segmentation in Python** - DataCamp
4906a9766f3355b5be0de117b2d449a93c9a92b1

 **Machine Learning for Marketing in Python** - DataCamp
5c37eef6fade2e7b02ddedf28a31e4c6a1190fdc

 **Advanced SQL** - Kaggle
gabrielardochoi/advanced-sql

 **AWS Machine Learning Foundations** - Udacity
9SMEGPP5

 **Data Scientist with Python Track** - DataCamp

173f940cafed51d38aa01338f3bc020f4bca2278



Statistics Fundamentals with Python Track - DataCamp

709822c9c90fb886273493420773993d1a1592bf



Databricks Lakehouse Fundamentals - Databricks

7e91d96d-fa38-48f1-9181-9991ad25b7b0



AWS Certified Machine Learning – Specialty - Amazon Web Services (AWS)

2207e8e0-ecaf-445b-8e14-1b9f11e7d6e6

Skills

MLOps • Git • Python (Programming Language) • Deep Learning • Natural Language Processing (NLP)
• Computer Vision • Amazon Web Services (AWS) • Data Science • SQL • Machine Learning