LUCAS MORAES

I am a data specialist that works both with data science and analytics. Having worked in very different environments gave me keen communication skills directed to both technical and non technical people, an asset I am valued for alongside my creativity. As a geneticist/bioinformatician by formation, I have been using data science tools for over a decade for academic, governmental and industry projects.

💼 RECENT PROFESSIONAL EXPERIENCE

Present | 2022

Senior Data Analyst
PicPay

- Data analysis and statistical support for the behavioral segmentation model (GMM) of customers.
- Data analysis, experimental design, data wrangling and experimentation monitoring for A/B testing of features in the suggestions session of the app home page.
- Data analysis for integrity check of models in production. Machine learning modeling for client propension studies.
- Analytical pipelines using GitHub + Databricks + Airflow for the creation of custom on demand tables in the data lake.

2022 | 2021

Data Scientist
Melhor envio

- End-to-end development of a machine learning client segmentation model for the company (K-prototypes).
- Statistical analyuses of client data for real time monitoring of user activity to aid decision making.
- Conversion of arbitrary business metrics to robust indicators using statistical techniques (e.g. bootstrapping and hypothesis testing).
- Dashboard modeling and development using Looker/LookML for technical and non technical audiences.

2020 | 2018

Independent consultant (Data Science & Analytics)
Freelancer

- Data compilation, wrangling, EDA and modeling for scientific reports and/or research projects.
- Experimental design and hypothesis formulation/testing.
- ML modeling for predictive, correlational and clustering analyses.

Contact & Portfolio

- 🏠 lucasmoraes.io
- ✉ moraessaur@gmail.com
- ⓰ GitHub: @moraessaur
- 🔗 LinkedIn

🎓 EDUCATION

2018
|
2016

MsC, Genetics

Rio de Janeiro Federal University

- Hierarchical clustering (unsupervised machine learning) and dendrogram analyses for evolutionary distinct lineages through the integration of biological, geographical and molecular unstructured data.
- Advisor: Carlos Guerra Schrago.

2012
|
2007

BsC, Genetics

Rio de Janeiro Federal University

- Phylogenetic and topological estimation of cetaceans using bayesian and maximum likelihood methods for hierarchical clustering parametrization.
- Publication: Phylogenetic Status and Timescale for the Diversification of Steno and Sotalia Dolphins. PLOS ONE. https://doi.org/10.1371/journal .pone.0028297
- Advisor: Carlos Guerra Schrago.

📕 ABOUT ME

Skills & Stack

- My area of expertise is tilted towards statistical modeling (Machine Learning included) to describe and understand client behavior, development of predictive models (e.g. churn or propension) and client segmentation using unsupervised ML. I have experience developing ad hoc analyses but also models aimed for production.
- Statistical modeling, data analysis and hypothesis testing are considered basic knowledge in the area I graduated from. This fact allowed me to transit between different technical areas seamlessly (such as machine learning, A/B testing and data visualization).
- R is my language of choice, but I also have proficiency with Python, SQL and Spark (PySpark/Sparklyr). I work comfortably with git, github and Linux, besides also having experience working in a Data Bricks environment (and with notebooks in general) and AWS (e.g. Redshift and S3).
- I have worked in several research projects with people from a wide array of backgrounds and seniority levels, some of which from end-to-end. I have been well trained since an undergrad to explain technical subjects to non technical audiences.

Technical knowledge

─────────────────────

®️ R

🐍 Python

🗄 SQL

Spark

📊 Statistics

🤖 Machine Learning

📈 Data Viz

👤 Fluent english

👥 Soft Skills