Data Engineer

Geneva, Switzerland
david.munoztord@mailbox.org
david-munoztord.com
munozt0
david-munoz-tord

Skills

- **R** 9 yrs.
- **Python** 5 yrs.
- **SQL** 5 yrs.
- **JavaScript** 3 yrs.
- **Bash** 7 yr.

Libraries

- tidyverse, data.table, dplyr
- TensorFlow, Keras, PyTorch, Scikit-learn
- Rstan, brms, JAGS
- H2O, XGBoost, LightGBM
- tidymodels, caret, mlr
- Sparklyr, PySpark

Tools/Frameworks

- Git, Terminal
- Docker, S3, EC2
- firehose, Athena, Glue
- lambda, Sagemaker, ECR
- ML OPS, CI/CD
- ETL, NIFI

Education

Ph.D. Candidacy in Computational Neuroscience

03/2020 - 01/2022
Campus Biotech - Switzerland

Summary

Dynamic and results-driven Data Engineer with a strong background in neuroscience, statistics, and data science. Experienced in developing and deploying comprehensive data solutions, collaborating with cross-functional teams, and leading innovative projects. Proficient in a wide range of programming languages and machine learning frameworks. Dedicated to continuous learning and staying at the forefront of technology trends.

Experience

Data Engineer — **10/2022 - Present**
Freelancer

– Led the development and launch of a scientific project in collaboration with EPFL, transitioning initial scripts into an accessible software to clients via the cloud.

– Offered technical consulting to a biotech start up, translating academic research into actionable solutions, while also advising on machine learning implementations.

– Within an NGO, collaborated with cross-functional teams to develop and implement data reporting infrastucture and implemented process automation best practices.

Data Engineer — **04/2022 - 01/2024**
FlowBank

– Contributed as a key member of the data migration to AWS cloud, ensuring seamless transfer, and scalability.

– Conducted in-depth analysis to classify customers using Bayesian inference, revealing distinct segments and patterns for data-driven decision-making and targeted strategies.

– Engineered and deployed a sophisticated fraud detection system, leveraging neural networks and support vector machines to accurately identify fraudulent activities.

– Developed new internal tools for data analysts in the form of R packages, and provided training and support for their use.

– Designed and implemented a datawarehouse solution for efficient data processing and orchestration, contributing to streamlined data engineering workflows.

– Led the development of a comprehensive ML-Ops infrastructure, leveraging AWS cloud infrastucture for seamless deployment and scalability.

– Designed and maintained a data visualization dashboard in R Shiny, which allowed for easy monitoring of key performance indicators.

Lecturer in Biotatistics — **09/2020 - 01/2022**
University of Geneva

– Conducted lectures on statistics for the graduate program and mentored students in their research projects, providing guidance on statistical analysis and data visualization.

– Created and shared freely available data workshops in R and Python.

PhD Candidate in Computational Neuroscience — **03/2020 - 01/2022**
Campus Biotech

– Processed terabytes of data for a large scale of longitudinal clinical trials (EEG/fMRI signal and blood samples).

– Used various statistical methods to analyze large scale datasets: Hierarchical Bayesian modeling, Multivariate pattern analysis, and model-free reinforcement learning.

– Demonstrated commitment to medical data management protocols compliance and proficiency in grant acquisition.

– Collaborated with Caltech data engineers to implement ICA-based denoiseification for better identification of activation and artifact components in fMRI.

Internship in Computational Neuroscience — **03/2019 - 03/2020**
Geneva Neurocenter

– Developed and implemented computational models to study neural system mechanisms and dynamics.

Topic: Computational modeling of addiction using neuroimaging data

MSc in Neuroscience

09/2018 - 03/2020
University of Geneva - Switzerland
Thesis: Denoising of high resolution fMRI signal using tICA

Complementary Studies in Computer Science

09/2017 - 06/2018
Smith College - USA
Capstone: Machine Learning for neuroimaging data

BSc in Cognitive Science

09/2014 - 06/2017
University of Geneva - Switzerland
Thesis: Early detection of ASD event related potentials

Languages

French Native

Spanish Native

English Fluent

Japanese Beginner

Technical Skills

- Data Sctuctures and Algorithms
- Data Mining
- Machine Learning
- Linux System Administration
- Visualization for Scientific Data
- Spatial Analysis
- Bayesian Statitics

Extracurricular

- Top 2% on StackOverflow (2024)
- Top 25% on GitHub (2023)
- Global Fellow scholarship (2017)
- Co-founder of Go-Fast, a socially responsible Bike Messenger cooperative in Geneva.

Projects

See my github profile for a more comprehensive list of open source projects.

Co-President — **We Data**
R, Python, Rust, JavaScript

– We Data is an organization that shares knowledge on code in data science (Blog, YouTube channel), give workshops about statistics, and do coding demonstrations. We also organize the R-Lunches.

Maintainer — **echarts4r**
R, JavaScript

– Collaborated on the development of echarts4r, a R package that expands the capabilities of R-based interactive visualizations with a powerful rendering engine, while enhancing flexibility and ease of use.

Collaborator — **hBayesDM**
STAN, R, Python

– Collaborated on the Hierarchical Bayesian modeling of Decision-Making tasks library (hBayesDM). Built Q-learning algorithm for probabilistic selection task.

Maintainer — **Firebase**
R, JavaScript

– Contributed extensively to the firebase integration package for R, empowering the Shiny user community by enabling user authentication and secure file storage through Firebase Storage.

Creator — **DbVieweR**
R, SQL

– Created a shiny app (DbVieweR) simulating a database management system featuring functions like login authentication, save/create/delete tables, add/rename columns, using either a PostgreSQL or SQLite back-end.

Collaborator — **reprtree**
R

– Integrated the caret ensemble for reptree (implementation of representative trees from ensembles of tree-based machines).

Collaborator — **3dLMEr**
BASH, R

– Collaborated on the AFNI's functions for 3Dimensional Linear Mixed-Effects Regression (3dLMEr). Fixed residuals output image by adding bottom tolerance.

Publications

See my publications pages for a comprehensive list my scientific publications.

References

Pierre Saouter, **Head of Data Science at FlowBank**

John Coene, **Founder and Lead developer at Opifex**

Jorge Figueiredo, **Head of Information Security at FlowBank**

Jose de Abreu Nunes, **Professor of Biostatistics at the University of Geneva**

Eva R. Pool, **Senior Researcher at Campus Biotech**