Raziel Amador Ríos

Barcelona/Spain | +34 644 847 924

razielar@gmail.com

[GitHub](#) | [LinkedIn](#)

Professional aims

Ph.D. Data Scientist with 5 years of experience driven by intellectual curiosity, continuing-learning, and passion for generating knowledge from data and being able to communicate a compelling story. Proficient in data mining and data analyses, statistical and machine learning techniques, cloud-computing, and report the results through visualizations. I possess strong problem-solving skills, and excellent communication abilities. I am looking to apply my analytical skills, implement AI/ML models, and bring Big Data solutions (Apache Spark).

Experience

Nennisiwok AI Lab   *2022-Present/ Barcelona*

SENIOR DATA SCIENTIST

- TIME SERIES MODELS: FORECASTING, ANOMALITY DETECTION | MODELS: LIGHTGBM, PROPHET, LSTM
- CHURN PREDICTION: BINARY CLASSIFICATION & SURVIVAL ANALYSIS | MODELS: CATBOOST, COX HAZARDS MODEL
- NLP: SENTIMENT ANALYSIS PROJECT & WORD EMBEDDINGS | MODELS: BERT, GPT2
- RECOMMENDER SYSTEMS: CONTENT RECOMMENDATION | MODELS: ALTERNATING LEAST SQUARE
- BIG DATA & ETL: APACHE SPARK, AZURE, DATABRICKS, AIRFLOW, SPARK ML MODELS, DELTA LAKE, SQL.

Centre for Genomic Research   *2017-2021 / Barcelona*

DOCTORAL RESEARCHER

- MULTICLASS CLASSIFICATION: IMBALANCE DATASETS | MODELS: CATBOOST
- UNSUPERVISED MODELS: PCA, T-SNE, UMAP, AUTOENCODERS.
- EXPLAINABLE AI (XAI): SHAP FRAMEWORK & INTERPRETABLE MODEL-AGNOSTIC FRAMEWORKS
- STATISTICAL LEARNING: BAYESIAN MODELS, POISSON REGRESSION.

SIRIS Academic   *2019-2020 / Barcelona*

DATA ANALYST

- LA CAIXA PROJECT: WEB CRAWLING, DATA ANALYSES, VISUALIZATIONS, AND DASHBOARDS

Centre for Genomic Research   *2016-2017 / Barcelona*

MASTER STUDENT

- REGRESSION MODELS: RESEARCH WORKING WITH CELL PROPORTION ESTIMATION | MODELS: SVM & XGBOOST

Education

University of Barcelona   *2018-2021*

- PhD IN BIOINFORMATICS AND DATA SCIENCE. *Grade*: **Cum Laude**

University of Barcelona   *2016-2017*

- M.S. IN BIOINFORMATICS

Tecnológico de Monterrey   *2010-2014*

- B.S. IN BIOTECHNOLOGY ENGINEERING

Skills

Languages	Spanish: ***Native***	English: ***Full professional proficiency*** (C1)
Programming Languages	Python & PySpark	R	SQL	Linux/Bash scripting
Frameworks & Libraries	Scikit-learn	PyTorch & TensorFlow	Matplotlib & Seaborn	Tidyverse & ggplot
Data & Statistical skills	Bayesian statistics	Big Data analysis	Survival analysis	Lineal & logistic regression
Software Development	Git	GitHub-Actions	Docker	High Performance Computing (HPC)
Cloud and databases	Apache Spark	Databricks	Azure & AWS	DataLake, Lakehouse (Delta tables)
Visualization tools	Plotly	Shiny	HTML & CSS	LaTeX	Markdown

ML & AI skills

Supervised Learning	Gradient Boosting: XGBoost, LightGBM, & CatBoost | Random-forest | SVM | kNN
Unsupervised Learning	Clustering: K-means & Gaussian mixture models | PCA, t-SNE & UMAP | Autoencoders
Deep Learning	Feed-Forward NN | CNN | GRU | LSTM
Forecasting/Time-series	ARIMA, SARIMAX & VAR | Prophet | LightGMB | LSTM | Time-series Clustering
Recommender systems	Alternating Least Square (ALS)
NLP	Transformers: BERT & GPT | Sentimental analysis | Word-embedding
MLOps	MLflow | ML Pipelines: Apache-Spark & Scikit-learn | Model Registry | Airflow

Certifications

- Specialized Models: Time Series and Survival Analysis. *Issued Feb 2023.* Coursera certification.
- Music Recommender System Using PySpark. *Issued Sep 2022.* Coursera certification.
- Building Machine Learning Pipelines in PySpark MLlib. *Issued Jul 2022.* Coursera certification.
- Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning. *Issued Apr 2019.* Coursera certification.

Honors & Awards

- EMERALD evaluator 2022. *Evaluator of international PhD candidates for medical doctors.* More information here.
- Barcelona Citython 2019. *Winner of the Comprehensive Cities category.* Using Deep Q-learning to propose a traffic and pedestrian mobility solution. More information here.
- Accenture Digital Healthcare Hackathon 2019. *Finalist (4th place).* Survival analysis in melanoma patients: developed a XGBoost algorithm to calculate patient survival probabilities.
- Barcelona Citython 2018. *Winner of the CISCO tech prize.* Anonymously count people crowds through Deep learning. More information here.
- Accenture Digital Healthcare Hackathon 2018. *Finalist (4th place).* Develop a random forest classifier to predict which patients with neutropenia will develop bacteremia.

Scientific Publications

- **Day-night and seasonal variation of human gene expression across tissues**. PLOS BIOLOGY, 2023. My contributions: Statistical analyses of large-scale gene expression data.
- **Genomic and functional conservation of lncRNAs: lessons from flies**. Mammalian Genome, 2022. My contributions: Transcriptomic and comparative analysis of the human, mouse and fruit fly genomes.
- **Unravelling the Role of Long Noncoding RNAs in the Context of Cell-growth and Regeneration**. PhD-Thesis, 2022. My contributions: Built an XGBoost classifier based on high-throughput CRISPRi functional screen data in seven human cell lines, as well as, cell-specific ENCODE data.
- **The effects of death and post-mortem cold ischemia on human tissue transcriptomes**. Nature communications, 2018. My contributions: building and training a support vector machine (SVM) model to infer cellular composition.