Jefferson Conza

Quito, Ecuador (GMT -5)
Phone number: +593 992701821, +51 953219234
E-mail: jeffconza3@gmail.com
LinkedIn: **jefferson-conza-46779b59/**
GitHub: **JeffersonConza**

Data Analyst | Machine Learning Engineer

Machine Learning Engineer with a background in Mathematics (BSc), and over 2 years of experience in developing models, deploying computer vision solutions, and building LLM-powered tools. Core strength in Python, R, SQL, Stata, and containerized applications such as Docker. Also have experience with TensorFlow, PyTorch, and NumPy. Seeking a challenging and innovative Machine Learning Engineer role where can apply and further develop my skills.
Eager to contribute to data-driven projects focused on solving complex problems.

Work Experience

Data Analyst at Ecuadorian Development Research Lab (LIDE) Jan 2025 to Oct 2025

- Refactored and generalized R algorithms for Synthetic Control Methods, enhancing code reproducibility and enabling the model's application across distinct datasets with varying temporal parameters (year).
- Feature engineered data integrating external economic indicators (GDP, Population) to resolve missing values and conducting Exploratory Data Analysis (EDA) to validate data quality.
- Implemented statistical inference models using R and Stata to quantify the impact of public policies in Ecuador, directly supporting data driven decision making processes.
- Documented research findings in technical reports, improving knowledge dissemination in Synthetic control (modeling for comparison of results).

Main Technologies: R, STATA, Excel, python

Data Analyst at Belozfi (YC W22) (Seasonal) Aug 2025 to Sep 2025

- Developed and refined machine learning models for real-time fraud detection using Python.
- Designed and proposed data driven campaign models based on SQL analysis to maximize conversion rates.
- Managed and optimized Search Engine Optimization (SEO) publishing efforts through hands-on use of Google Search Console. This included monitoring website performance, identifying technical SEO issues, analyzing search queries to inform content strategy, and ensuring content visibility and organic traffic growth.
- Presented weekly reports (1-2 pages) to show project milestones, SEO metrics, and key performance indicators, ensuring stakeholder alignment and transparency on progress.

Main Technologies: Python, SQL, Slack, MetaBase, Google Search Console

Machine Learning Fellow at Anyone AI (Part-time) Jan 2025 to Jun 2025

- Developed sentiment analysis and word embedding models to classify movie reviews, leveraging BoW, TF-IDF, and deep learning techniques to detect polarity in unstructured text.

- Predicted home credit repayment capability using a dataset of 350K+ transactions, achieving ROC AUC >0.72 with DecisionTree, XGBoost, and LightGBM models.
- Built a data pipeline and visualizations for Latam e-commerce analysis (2016–2018), uncovering trends in revenue, product categories, and delivery delays across regions.
- Deployed an image classification model with 82% accuracy to identify car make and model from images; containerized as an API-based web service using Docker and AWS.

Main technologies: Python, Numpy, Pandas, Scikit-learn, Keras, PyTorch, Docker, AWS, SQL, Seaborn, Matplotlib

Data Annotator at TALOV April 2022 - August 2022

- Collected and curated a diverse dataset of currency images, ensuring representation of various conditions (e.g., well-preserved, damaged, folded, inked) to enhance object detection accuracy for visually impaired users.
- Conducted image augmentation on various datasets of 100+ images (for currency), generating rotated variations (90°, 180°, and 270°) to improve model generalization and robustness.
- Updated the dataset with newly released currency, maintaining model relevance and accuracy in real-world scenarios.
- Contributed to a neural network achieving over 90% recognition accuracy in currency detection, improving accessibility for visually impaired individuals.

Main Technologies: *Roboflow, Python, Excel*

Projects

Isotropic Reconstruction of Bio-Imaging Data (Confidential Research) Aug 2025 - Present

- Developing a *CycleGAN* architecture in PyTorch to perform unpaired image-to-image translation, mapping axial (xz) source domains to high-fidelity lateral (xy) targets to correct anisotropic distortion.
- Engineered a high performance data ingestion pipeline to process *120,000+ image patches* (128x128 resolution) across 12 diverse datasets (Nuclei, Membranes, Bile Canaliculi), handling complex **.npz** tensor extraction.
- Documenting neural network architecture and experimentation results in LaTeX, establishing a baseline for the research of the engineering team.

Main Technologies: Python, PyTorch, GitHub, NumPy, LaTeX

Energy Forecast Hub | *View on GitHub*

- Architected an end-to-end forecasting system serving predictions via a FastAPI backend and Streamlit dashboard, orchestrated with Docker Compose for reproducible deployment.
- Implemented and benchmarked regression models (*Linear Regression*, *Random Forest*, and *XGBoost*) in both Python and R, aimed to compare, selecting Random Forest as the production model based on optimal RMSE performance.
- Refactored a legacy codebase into a modular package structure, integrating automated data pipelines for historical training and multi-day forecasting.

Main Technologies: Python, R, Docker, FastAPI, Streamlit, XGBoost, Scikit-learn

Pneumonia Detection with Transformers | *View on GitHub*

- Implemented the *CheX-DS* architecture (*DenseNet121 + Swin Transformer*) in PyTorch, utilizing a Weighted Asymmetric Loss to effectively handle class imbalance in medical datasets.
- Achieved 99% Recall (Sensitivity), outperforming a ResNet50 benchmark by reducing Test Loss (0.29 vs

1.03) and minimizing critical false negatives for patient safety.
- Deployed the diagnostic model via an interactive Streamlit dashboard, containerized with Docker to ensure reproducible inference across diverse computing environments.

Main Technologies: Python, PyTorch, Swin Transformer, Docker, Streamlit.

Skills

Tech Skills: Python, R, Julia, Kotlin, SQL, Stata, Bash
Frameworks & Libraries: PyTorch, TensorFlow, Keras, Scikit-learn, Pandas, NumPy
Agile Methodologies: Scrum
Tools & DevOps: Docker, Git/GitHub, VS Code, PyCharm, Antigravity CLI, Wolfram Mathematica, LaTeX
Languages: Spanish, native; English, C1

Education

Yachay Tech University Feb 2019 - Expected Graduation: Jun 2027
BSc. Mathematics

Anyone AI Jan 2025 - Jun 2025
Machine Learning Engineering Program

AWS Cloud Institute Oct 2025 - Jun 2026
AWS Cloud Developer Certificate

Courses

DataCamp | Llama Fundamentals February 2025

DataCamp | Deep Learning in Python (PyTorch) January 2025

NVIDIA | Fundamentals of Accelerated Computing with CUDA Python September 2024

Platzi | Visualización y Storytelling con Latinometrics September 2024