

Mario Arturo Paulin Vega

Union 28, Miguel Hidalgo, CDMX, Mexico

mpaulinv@gmail.com

linkedin.com/in/mario-paulin-vega

Portfolio: github.com/mpaulinv

Summary

Data Scientist with eight years experience delivering ML solutions across financial services, economic research, and consulting. Led the validation and deployment of production models across Mexico, LATAM, and North America, specializing in predictive modeling, anomaly detection, and statistical analysis using Python, SQL, and big data tools. Built and managed data science teams while collaborating with engineering and business partners to drive operational efficiency. Passionate about applying analytics solutions to solve complex business problems.

Education

MSc in Data Science , The University of Texas at Austin	Jan 2022 – Dec 2024
MSc in Economics , The University of Texas at Austin	Sep 2012 – Jun 2014
Master in Economic Theory , ITAM	Aug 2010 – May 2012
Bachelor in Economics , Finance Track, With Honors, ITAM	Aug 2004 – Dec 2008

Work Experience

Team Lead, Anti-Money Laundering Division

CitiBanamex	Mexico City	Jun 2020 – May 2024
-------------	-------------	---------------------

- Managed a team of data scientists in validating a regulatory model inventory of around 100 models across Mexico and the US, ensuring full compliance with regulatory timelines and achieving zero audit and regulatory findings.
- Led code, data reconciliation, and algorithm reviews, uncovering critical implementation issues resulting in millions of dollars in previously untagged operations, enabling corrective action and improved model performance.
- Built and led 4-person data science team from inception, mentoring junior scientists with a 50% promotion rate to leadership roles.
- Presented an initiative to global senior compliance management to reduce by 15% time spent on Manager Control Assessment across the Bank.

Senior Statistician, Anti-Money Laundering Division

CitiBanamex	Mexico City	Jun 2017 – May 2020
-------------	-------------	---------------------

- Built complete life-cycle validation pipelines and produced comprehensive reports for predictive and rule-based models, ensuring full regulatory compliance and contributing to zero findings from Internal Audit and external regulators.
- Executed end-to-end model validation, including data extraction, code review, replication, and quality assessments in production and test environments; applied advanced statistical techniques such as Logistic Regression, Decision Trees, Random Forest, Lasso, and K-means clustering to analyze transactional and client profile data.
- Executed a global project automating data input stability assessment, reducing time spent by 50%.

Independent Consultant

Freelance

Queretaro

Aug 2016 – May 2017

- Developed econometric models to analyze labor transition probabilities in Colombia (IDB-funded project).
- Automated data workflows and dashboard pipelines for structural transformation analysis (World Bank-funded project).

Research and Teaching Assistant

The University of Texas at Austin & CIE

Austin and Mexico City

2011 – 2016

- Taught over 100 hours of undergraduate and master level economics courses.
- Analyzed export dynamics using firm-level panel data with econometric models.

Impact Evaluation Analyst

CONEVAL

Mexico City

2009 – 2010

- Evaluated social development programs using DiD, RDD, PSM, and panel data to assess program effectiveness. Coordinated a user satisfaction survey project, participating in survey design, data collection, and statistical analysis with an external consultant

Volunteering

Data Quality Analyst

Statistics Without Borders

Jan 2025 – Present

- Supported development of a white paper on AI use in NGOs, ensuring data consistency.
- Contributed to an NLP and automation project streamlining sustainability reporting.

Certifications

- LangChain for LLM Application Development, DeepLearning.AI (2025)
- Harnessing Language Models: Your Path to NLP Expert, Barcelona (2024)
- Certified Anti-Money Laundering Specialist (ACAMS) (2021)

Technical Skills

Programming: Python (Pandas, NumPy, Scikit-learn, PyTorch), R, SAS, Stata

Big Data and Cloud: SQL, Hive, Spark, Hadoop, Google Cloud Platform

Visualization: Matplotlib, Seaborn, ggplot2, Power BI

Machine Learning: Regression, Clustering, Decision Trees, Random Forest, Neural Networks, NLP, Feature Engineering

Statistics: Hypothesis Testing, Causal Inference, Econometric Methods

Languages

Spanish (Native), English (Fluent), Italian (Beginner)