

Sebastian Mateus

[linkedin.com/in/sebastianmateusperdomo](https://www.linkedin.com/in/sebastianmateusperdomo)

HONORS AND MENTIONS

Data Science & Machine Learning Divulgator Autonomous	Bucaramanga, COL February 2022 - Present
● TEDx Speaker: “IA, Guía Rápida Para Cambiar El Mundo” (AI, Quick Guide On Changing The World) ● Scientific Blogger: Currently writing about data science, artificial intelligence, and science on Medium	
Kaggle Competitor Autonomous	Bucaramanga, COL February 2025 - Present
● Kaggle Expert, competed in 18 hackathons, ensuring top 20% in 8 of those, achieving 1 Silver Medal	
International Hackathon Winner Baubap’s Machine Learning Hackathon & Baubap’s AI Hackathon 2.0	CDMX, MEX August 2023
● Won first place in the neural network’s category of the Mexican fintech Baubap, creating an algorithm fed by more than 1K features in the form of tabular data that allowed determining fraudulent transactions with +90% accuracy.	

PROFESSIONAL EXPERIENCE

Nequi (Grupo Bancolombia). <i>Senior AI/ML Engineer</i>	Colombia (Remote) Nov 2025 – Present
● Own the data and deployment infrastructure for production ML at a leading LatAm neobank (Grupo Bancolombia), on AWS and Databricks; ensure reliable ML pipelines, feature engineering, and data quality across the credit-risk and origination model stack in a regulated financial environment. ● Built a foundational model from scratch, pre-trained on the full customer database (transaction and behavioral event sequences), to serve as a shared representation layer that materially improves origination-model performance and consolidates multiple downstream ML products into a single reusable backbone. ● Designed, developed and deployed a wide range of models end-to-end—LLM/RAG systems, gradient-boosting and classical ML, and graph-based models—owning the full lifecycle from experimentation through deployment, CI/CD, and production monitoring.	
Factored. <i>Competitive ML Team Lead & Founder</i>	Palo Alto, CA (Remote) January 2025 - Jan 2026
● Founded and currently Leading the ML Competitive team inside of Factored, grew the team from 1 to 13 engineers in 10 months and also participated in 5 international challenges in that time, getting top 20% in all of the competitions we’ve participated in so far. Scored top 3% in the ARC prize 2025	
<i>Machine Learning Engineer (L3)</i>	June 2025 - Jan 2026
● Architected a high-scale data ingestion platform using Dockerized Scrapy containers and AWS Batch, orchestrating the collection of school-related data from hundreds of sources; implemented event-driven Lambda functions to normalize heterogeneous data into a universal schema, reducing ingestion time by 40%. ● Developed an end-to-end MLOps pipeline for a GenAI city-guide product: RAG workflows with hybrid retrieval (dense + BM25) and cross-encoder reranking via LangChain, LLM-as-judge evaluation pipelines (faithfulness, Recall@5), AWS Glue for feature aggregation, and Terraform (IaC) for 100% Dev/Prod parity. ● Optimized production model reliability by implementing automated Data and Concept Drift detection using CloudWatch and Langfuse (LLM tracing: latency, token cost, error rates per request); achieved a 15% improvement in Precision@5 through systematic offline retraining and Champion-Challenger validation cycles.	
<i>Machine Learning Engineer (L2)</i>	June 2024 - June 2025
● Designed and deployed end-to-end ML pipelines—incorporating web-scraped and third-party data ingestion, preprocessing workflows, and generative-AI content creation at scale (LLM APIs, structured prompt engineering, automated quality validation)—to accelerate model development and deployment. ● Architected on Python-based recommender systems to model user behavior and preference segmentation, enabling more precise content recommendations and increased user engagement. ● Leveraged ML-driven geospatial tools to validate and correct location data across all company datasets, ensuring consistent geographic accuracy for analytics and decision-making.	
Intouch CX. <i>Senior Data Scientist</i>	Winnipeg, MB (Remote) July 2023 – April 2024
● Employed Python to analyze large amounts of raw data, apply cleansing processes and pre-trained transformer/NLP models (text classification, embeddings) to extract insights and guide data-driven decisions with stakeholders enhancing business strategies. ● Led Python-based machine learning projects to predict service demand behavior, contributing to workforce restructuring planning.	

- Developed Dashboards powered by ETL's for tracking KPI's and direct interaction with the company's customers as a method of engaging sales.

Robin Food.
Analytics Analyst

Bogotá, CO-CUN
April 2022 - August 2023

- Built and implemented NLP algorithms for the user happiness area, used to study, classify and send to a MySQL server about 10K custom comments per month, significantly reducing the work of UH analysts.
- Built and implemented a user clustering algorithm used to classify nearly 1M of the company's customers, allowing the marketing and growth team to perform personalized communication and increase user retention.
- Used python to develop and automate more than 10 ETL processes that fed Tableau dashboards for all company KPIs.

Immersion Design.
Product Owner for Data Science

Bucaramanga, CO-SAN
January 2021 - April 2022

- Developed metrics and consultation processes for performance monitoring of developer teams
- Effectively prioritized and managed product backlogs, ensuring that the most valuable features are delivered in a timely manner.
- Collaborated closely with cross-functional teams, including developers, designers, and stakeholders, to define product vision, goals, and requirements while implementing agile methodology.

Vision Puntual SAS.
Chief Data Scientist & Co - Founder

Bucaramanga, CO-SAN
April 2019 - December 2021

- Implement a churn prediction model for a database with over 10K customers to identify customers who are at risk of leaving and implement targeted retention strategies.
- Implement dynamic pricing models that consider factors such as customer demographics, purchasing behavior, and market trends to optimize pricing strategies for different eyeglass frames and lenses.
- Use machine learning to detect and categorize imperfections, reducing the likelihood of faulty products reaching customers.

Metric Analytics.
Chief Data Scientist

Bucaramanga, CO-SAN
April 2018 - December 2020

- Developed a ML algorithm capable of predicting the behavior of financial products in the stock market used in the management of portfolios of millions of COP
- Successfully led and managed a high-performing data science team fostering a collaborative and innovative work environment.
- Developed ETL and DataViz processes for OKR's follow-up

SKILLS

- **Programming & Software Engineering**
Languages: Python (Expert), SQL (Expert), Scala, R, HTML/CSS (Frontend Very basics).
Tools: Docker, Terraform (IaC), Git/GitHub, Unit Testing (Pytest), Linux/Bash, REST APIs (FastAPI/Flask).
Concepts: Object-Oriented Programming (OOP), Data Structures & Algorithms, CI/CD Pipelines.
- **MLOps & AI Infrastructure**
Frameworks: MLflow (Lifecycle & Registry), Vertex AI (GCP), Databricks, Azure Machine Learning, SageMaker.
Model Operations: Model & Data Drift Detection, Champion-Challenger Validation, Hyperparameter Tuning (Hyperopt), Model Serialization (Pickle, Joblib, ONNX).
AI/ML: LLMs & Gen-AI (Claude/OpenAI/Gemini APIs), RAG & Vector DBs (Pinecone, ChromaDB), LangChain/LangGraph, Langfuse (LLM Observability), Agentic Systems & MCP, Prompt Engineering, LLM Evaluation (LLM-as-judge), Recommender Systems, Scikit-learn, PySpark MLlib.
- **Data Engineering & Big Data**
Distributed Compute: Apache Spark (PySpark), Hadoop.
Storage & Databases: BigQuery (GCP), Delta Lake (Lakehouse Architecture), Azure Blob Storage/S3, MySQL, MongoDB, PostgreSQL.
Pipelines: dbt, Airflow, Azure Data Factory, AWS Batch, Scrappy (Distributed Scraping), ETL/ELT workflow design.
- **Analytics & Visualization**
Tools: Tableau, Power BI, ElasticStack (ELK), Product Analytics (Amplitude, GA, Clevertap, Mixpanel).
Libraries: Pandas, NumPy, Matplotlib, Seaborn.
- **Spoken Languages:** English, Spanish, Portuguese