Esteban Vilca

Data Scientist & Data Engineer

linkedin.com/in/estebanvz

SUMMARY OF QUALIFICATIONS

Data Scientist & Data Engineer with 7+ years in data and machine learning, including financial reporting, FX analytics and cryptocurrency modeling. Led a five-country migration with an estimated 80% reduction in pipeline costs (approximately USD 60,000 annually). Brings cloud delivery and peer-reviewed research to finance, fintech and cryptocurrency teams. PhD Candidate in Computer Science. Communicates in English, Spanish and Portuguese.

TECHNICAL SKILLS

Data: Python, SQL, Snowflake/dbt, PySpark, Apache Airflow
Cloud and delivery: AWS Glue/Athena, GCP BigQuery/Dataproc, Azure, Git/Docker
Machine learning: TensorFlow/PyTorch, scikit-learn, time-series modeling, FX analytics
AI tools and APIs: Binance API, OpenAI API/Batch API, Codex/GitHub Copilot

PROFESSIONAL EXPERIENCE

Data Engineer | Squadric

Sep 2025 – Present | Canada

Technology services. End client: Finning Canada, a Caterpillar equipment dealer.

- Led the five-country Databricks-to-Snowflake migration after a team restructuring, reducing pipeline costs by an estimated 80% (approximately USD 60,000 annually).
- Built and maintained approximately 25 production pipelines, with workloads from around 2 GB every 30 minutes to approximately 1 TB in weekly processing.
- Cut one pipeline runtime by approximately 25% (20 to 15 minutes); implemented reconciliation, automated QA and monitoring. Used Codex and GitHub Copilot for code development.

Data Scientist | Vigil

Aug 2024 – Sep 2025 | Canada

Financial reporting and foreign-exchange data operations.

- Automated reporting across three European countries, replacing Excel/email workflows and supporting approximately 10,000 financial operations per week.
- Built AWS Glue, Athena and Python pipelines for financial and FX data; analyzed trading profitability, execution quality and risk for Finance teams.
- Designed pipelines integrating the OpenAI API and Batch API; applied anonymization and governance controls to sensitive financial information.

Data Engineer | Amaris Consulting

Sep 2022 – Aug 2024 | Canada

Technology consulting. End client: Cencosud, a Latin American retail group.

- Led a multi-country retail migration processing approximately 1 TB on GCP, using PySpark, Dataproc, BigQuery and Airflow for data and ML pipelines.
- Validated the solution through a controlled store comparison that showed approximately 10% higher sales after deployment.

Data Scientist Instructor | Alura Latam

Oct 2021 – Sep 2023 | Brazil

Online technology education platform serving Latin America.

- Delivered four courses on neural networks and time-series analysis, reaching approximately 100 students with an average learner rating of approximately 4.5/5.

Data Scientist | Semantix Brasil

Sep 2021 – Sep 2022 | Brazil

Brazilian data and AI technology company.

- Built GCP streaming/batch and TensorFlow/PyTorch workflows for sensor and video data; deployed AI on Jetson and Raspberry Pi and used anonymization and face blurring to protect sensitive patient information.

Data Scientist | Eyachay Software and Education

Sep 2019 – Sep 2021 | Peru

Software and education business; financial analytics and ML projects.

- Developed financial ML and cryptocurrency time-series models using Python, scikit-learn, TensorFlow/Keras and LSTM, CNN and GRU architectures.
- Integrated the Binance API for market-data collection and automated long/short trading workflows; used Snowflake for financial and social-market datasets.
- Built AWS/GCP workflows and financial analytics dashboards with Metabase, Power BI and Tableau.

RESEARCH AND PUBLICATIONS

Bitcoin Transaction Networks for Portfolio Profitability

Procedia Computer Science | 2023 | Peer-reviewed

Led the research and manuscript writing on transaction-network features and deep learning for cryptocurrency portfolios.
DOI: 10.1016/j.procs.2023.08.192

Quipus: Classification Using Attribute Interactions

Incodtrin (IEEE) | 2020 | Peer-reviewed

Primary researcher and manuscript author; modeled attribute interactions for graph-based classification.
DOI: 10.1109/Incodtrin51881.2020.00026

Classification Using Betweenness Centrality

ENIAC (SBC) | 2020 | Peer-reviewed

Led research and manuscript writing; evaluated the classifier on nine datasets against nine established methods.
DOI: 10.5753/eniac.2020.12128

COVID-19 Dataset Characterization Using Complex Networks

arXiv preprint | 2020

Medical-image feature extraction and pattern analysis using complex networks and image processing.
arXiv:2009.13302

Twitter Reactions to Impeachment in Peru

arXiv preprint | 2020

Analysis of Twitter discussions and public reactions to impeachment in Peru.
arXiv:2010.04877

EDUCATION

PhD Candidate in Computer Science | Universidade de Sao Paulo | Sep 2021 – Present
Master of Science in Computer Science | Universidade de Sao Paulo | Mar 2019 – Sep 2021
Master of Management | Crandall University | May 2023 – Dec 2024
Bachelor's degree, Computer Science | Universidad Peruana de Ciencias Aplicadas | 2014 – 2018
ACP Data Science & Predictive Analytics for Business Professionals | UC Irvine | 2018
ACP Innovation Management & Entrepreneurship | UC Irvine | 2018

CERTIFICATIONS

SnowPro Associate: Platform Certification | Snowflake
Microsoft Certified: Azure Data Scientist Associate

PROFESSIONAL DEVELOPMENT

Ethereum Token Implementation Course: ERC-20
DP-100 Exam Prep: Designing and Implementing a Data Science Solution on Azure
Introduction to Kubernetes
School of Blockchain and Cryptocurrencies