

VALENTINA TORRES DA SILVA

vt2450@columbia.edu | linkedin.com/in/valentinasilva8 | github.com/valentinasilva8 | Portfolio

EDUCATION

Columbia University

MS in Data Science

Relevant Coursework: Agentic AI, Design & Analysis of Online Experiments, Algorithms for Data Science

New York, NY

Expected Dec 2027

Fairleigh Dickinson University

BS in Computer Science, Minor in Mathematics, GPA: 3.96/4.0

Big Data Analytics Concentration | NCAA Division I Tennis Student-Athlete

Teaneck, NJ

May 2026

WORK EXPERIENCE

Women in Data Science (WiDS) Worldwide

Remote

Machine Learning Engineer Intern

Sep 2025 - May 2026

- Developed the 2026 WiDS Global Datathon (4,500+ entrants, 21K+ Kaggle submissions), standardizing sparse wildfire perimeter data, identifying spatial target leakage, and reframing wildfire risk as censored survival analysis.
- Built propensity and fraud models for WiDS Elevate; improved fairness in propensity ranking, increasing the share of customers ages 55+ in the top 10% of predictions from 9.6% to 16.4% with a 2.3-point recall tradeoff.
- Developed a fraud model on 45K simulated transactions, fixing point-in-time leakage in crypto wallet risk scores to raise PR-AUC from 0.45 to 0.85 with 71% recall; translated model scores into approve, review, or block decisions to optimize P&L.
- Selected as 1 of 5 judges for a WiDS and Accenture Song university AI apprenticeship, evaluating technical and business solutions from 6 universities focused on equitable wildfire evacuation and economic resilience.
- Presented on structural bias and algorithmic accountability at the WiDS 2026 AI Impact Forum at Accenture's San Francisco Innovation Hub; delivered a technical talk on wildfire ML challenge design for Break Through Tech's Virtual AI Program.

ACADEMIC PROJECTS & EXPERIENCE

Columbia University Data Science Institute

Sep 2026 - Present

DSI Scholar Researcher

- Built a Python web-scraping pipeline with Playwright and BeautifulSoup to turn multi-decade university newspaper archives into a standardized corpus for research on ideological change.
- Deployed long-running scrapers on AWS EC2 with checkpointing and failure recovery, using S3 backups and a reusable site adapter to support continuous collection across heterogeneous newspaper platforms.

Fairleigh Dickinson University

Dec 2025 - Jun 2026

Researcher

- Developed a four-class river-ice classifier for a NOAA-funded project, combining VIIRS imagery with sub-pixel water fraction to reach 78.1% cross-validated accuracy on narrow Alaskan rivers.

Fairleigh Dickinson University, Database Systems Course

Jan 2026 - May 2026

Teaching Assistant

- Graded homework and exams for a class of more than 50 graduate and undergraduate students.

RAG Legal Contract Extraction | FDU × RSG Media/Rightsline

Aug 2024 - Nov 2024

- Built a RAG contract-extraction system for RSG Media using OpenAI embeddings and XML-enriched text to automate structured extraction from legal contracts, earning 5/5 client feedback.

LONG RUN | HSI Battle of the Brains, Best Technology Solution Winner | Team Leader

Oct 2024

- Led an 8-person team to create and pitch LONG RUN, a sustainable outdoor startup concept, building its business model, financial projections, MySQL database, and Next.js/React prototype in 16 hours; won Best Technology Solution and 3rd place (\$10K) before judges from eBay, Capital One, Dell, and KPMG.

AWARDS

eBay Innovation Scholarship (\$5,000) | 1 of 4 recipients

Oct 2024

SKILLS: Python, SQL, R, scikit-learn, XGBoost, PyTorch, TensorFlow, Hugging Face Transformers, FastAPI, A/B testing