Lucas Flores, Ph.D.

San Francisco, CA, United States
+1 (951) 545-3382 lucasmacrorieflores@gmail.com linkedin.com/in/lucas-m-flores/

Education

Aug 2015 – Dec 2021 **PhD & MS** – Particle Physics *University of Pennsylvania, Philadelphia, PA*
Sep 2010 – Jun 2015 **BS** – Physics & Applied Mathematics *UC Riverside, Riverside. CA*

Experience

Data Scientist @ *Albeado* – Santa Clara, CA [Remote] Oct 2023 – PRESENT

- A full stack Data Scientist in a small, fast-paced, and innovative startup that builds explainable data-driven products from the ground up
- Identified and sourced Bitcoin address labels via WalletExplorer REST API, combining/integrating this data with BigQuery's Bitcoin blockchain dataset
- Ideated an asymmetrical approach to a breadth first search based algorithm for identifying the shortest transaction path between two Bitcoin addresses, leading to an up to 30x reduction in search time
- Developed and implemented a suite of heuristics derived from the Bitcoin blockchain protocol, identifying Bitcoin transactions that used privacy services, reducing noise in down-the-line community detection as well as identifying risky entities
- Designed experiments, built the data pipeline, and developed predictive machine learning models for a healthcare insurance claims dataset based product, identifying medical procedures as predictive indicators for early intervention, mitigating costly spine, hip, and knee surgeries
- Meet regularly with stakeholders, presenting latest insights, distilling technical concepts, and extracting high level business questions and goals, translating them into actionable data-driven projects and studies
- Highly cross-functional, partnering with Product to build mock-ups for potential partners and regularly meeting with the company's core team to ideate solutions, discuss analytical results, develop technical and business approaches, formulate decision making
- Investigated Generative AI/LLM uses cases both internally and for forward-facing products

Research Assistant/Physicist @ *University of Pennsylvania* – Philadelphia, PA Jul 2015 – Dec 2021

- Investigated petabytes of data produced by the Large Hadron Collider physics experiment in search of theorized subatomic particles, furthering the knowledge of fundamental physics
  - Generalist in a highly collaborative 5-member analysis team
  - Researched, constructed, optimized, and implemented two new features used to form control, validation, and signal regions for robust statistical hypothesis tests
  - Processed big data sets utilizing cloud services and distributed computing software (HTCondor [TORQUE/SLURM analog], RUCIO [Hadoop analogue], and PanDA [Airflow analog])
  - Set exclusion limits at a 95% confidence level over a large parameter space scan for the existence of new fundamental particles via a profile likelihood ratio fit
  - Communicated technical methods and results to the broader physics community at two international conferences
- Performed model tuning, software development, and new user onboarding/mentorship in an expert level leadership role in a major division within the collaboration
  - Instructed, mentored, and guided 4 new technical users and developers of the codebase/analysis
  - Re-optimized 108 independent multivariate likelihood models designed to identify/classify electrons, This is integrated in nearly every analysis in the 5000+ member collaboration
    - Trained models on a 20% larger, most current, and most representative data set
    - Tuned 324 selection parameters (3 per model), achieving targeted performance benchmarks
    - Transitioned training models from a 25% simulated/75% real data hybrid to 100% real data
  - Investigated a new metric for tuning the electron classifier to retain desired signal and background rates for different particle detector environments
  - Re-tuned critical classifier parameter that created a gain in signal rate in 50% busier detector environments

Skills

Languages: Python, C++, BASH, SQL (SAP IQ DB, BigQuery), HTML/CSS, JavaScript
Tools: UNIX, Git, CI/CD, pandas, NumPy, scikit-learn, NetworkX, PyWhy/DoWhy/causal-learn/EconML Keras, TensorFlow, matplotlib, HTCondor, Docker, Amazon Web Services (AWS), JIRA, REST APIs
Skills: Fundamental research, experiment building, hypothesis testing, statistics, physics, machine learning, data visualization, data mining, web design, web scraping