

JEREMY THALLER

Senior full-stack data scientist with two MS degrees and experience productionalizing machine learning models in fintech and logistics

EXPERIENCE

VEHO | Last-mile delivery company powered by a dynamic driver marketplace

Senior Full-Stack Data Scientist, Marketplace | New York, NY (Remote) [May. 2023 – Present]

- Designed and led the implementation of a major overhaul to the route generation algorithm, coordinating cross-functional changes across engineering and operational scheduling teams; this initiative achieved record-breaking efficiency and cost savings upon launch, translating to 17 cent per package saving or \$2M in yearly savings
- Led the design and development of a distributed computing system for route simulations, creating an end-to-end pipeline that included ETL processes, scalable EC2-hosted simulations via ray, and integration of the results via redshift and DBT into a dashboard
- Implemented stateless supervised and unsupervised ML algorithms within the route-generating microservice
- Refactored external API calls to spin through a rate-limiting backoff loop, log metrics, and trigger Pager Duty alerts
- Engineered new dry ice routing requirements, defined new data contracts, and optimized parameters through simulations for a successful launch

CURRENT | Financial technology company with no-fee banking, payroll deposits, debit and credit cards

Data Scientist, Core-Banking / Product | New York, NY (Hybrid) [Aug. 2022 – May 2023]

- Achieved 95% accuracy in predicting long-term, user-level segmentation from 30d of user behavior; segments were assigned via k-means clustering and PCA, and the early indicators model was trained with Vertex AI and BigQuery ML; resulting model provided feedback for marketing spend optimization in 1/6 the time as previously required.
- Modeled different product gating criteria to forecast potential churn and revenue scenarios; implementation resulted in \$200k/month in cost savings, as estimated through causal impact analysis determined with SARIMA forecasting.
- Wrote DAGs and SQL queries to populate daily updating BQ tables powering Looker dashboards; later contributed to the DBT migration from Airflow

CELSIUS NETWORK | Centralized finance crypto company providing various financial services, formerly with \$30B AUM

Data Analyst, Growth and Marketing | New York, NY (Remote) [Sept. 2021 – Aug. 2022]

- Utilized supervised machine learning to identify potential accredited investors with 98.4% accuracy; expanded the number of high net worth prospects by 10x, delivering \$50B of potential platform growth
- Determined feature importance for third party prospecting via unsupervised machine learning (clustering)
- Designed, developed, and deployed an ML anomaly detection algorithm to alert fraud and platform violations
- Developed internal Python packages to standardize and expedite repeated SQL queries and data transformations
- Initiated data augmentation through user surveys; utilized qualitative and quantitative results—often with Bayesian statistics—to predict vectors of user and asset growth, as well as steer brand image and marketing strategies
- Developed and presented weekly research to executive stakeholders, steering strategic GTM initiatives

BROOKHAVEN NATIONAL LABORATORY | Structure and Dynamics of Applied Nanomaterials

MS Thesis Researcher in Deep Learning | Upton, NY [Feb. – Sept. 2021]

- Reduced simulation compute time by 50x by developing a new statistical-based methodology
- Utilized TensorFlow and convolutional neural networks to predict absorption spectra disorder, reducing required data by 90%
- Mapped the simulation-trained model onto experimental data via transfer learning

EDUCATION

LUDWIG MAXIMILIANS UNIVERSITÄT, Munich, Germany | 2019 – 2021

M.S. Materials Science and Engineering

ADAM MICKIEWICZ UNIVERSITY, Poznań, Poland | 2019 – 2021

M.S. Computational and Applied Physics

WILLIAMS COLLEGE, Williamstown, MA | 2015 – 2019

B.A. Physics, Honors | Sigma Xi Honor Society | Varsity Track & Field Captain

SKILLS AND TOOLS

Programming Languages (Years of Experience) | Python (5+), SQL (4+), Java (7), MATLAB (4), R (1), Julia (1)

Python Packages | Pandas, NumPy, Scikit-Learn, Numba, PyTorch, TensorFlow, Keras, PySpark, Regex, WandB, Dask

Data Visualization Software | Looker, Data Studio, Excel/GSheets, Mathematica, Jupyter Notebooks, WandB, Plotly

Data Engineering Tools | Snowflake, Anyscale (ray), Apache Airflow, Docker, PySpark, DBT, AWS