Email | jeremy@thaller.dev
GitHub | github.com/jthaller
Portfolio | jeremythaller.com

JEREMY THALLER

Brooklyn, NY

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

EXPERIENCE

VEHO | Last-mile delivery company powered by a dynamic driver marketplace
Senior Data Scientist, Marketplace | *New York, NY [May. 2022 – Present]*

- Revamped route generation algorithm, leveraging simulations to optimize variables and implemented changes that yield a substantial 15 cents per package cost reduction, translating to over $1M in annual savings.
- Implemented stateless supervised and unsupervised ML algorithms within the route-generating microservice
- Refactored external API calls to spin through a rate-limiting backoff loop and log success metrics in new relic

CURRENT | Financial technology company including no-fee debit banking, debit cards, early payroll deposits, and crypto trading
Data Scientist, Core-Banking / Product | *New York, NY [Aug. 2021 – May 2022]*

- 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 data studio dashboards; later contributed to the DBT migration from Airflow

CELSIUS NETWORK | Centralized finance crypto company providing comprehensive financial services, formerly with $30B AUM
Data Analyst, Growth and Marketing | *New York, NY [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 to predict absorption spectra disorder of Au nanoparticles, reducing data required by 90%
- Created and managed lab's GitHub organization; constructed example projects to demonstrate best dev practices

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 (3), 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, Apache Airflow, Docker, PySpark, DBT, AWS