jeremy@thaller.dev github.com/jthaller • jeremythaller.com

JEREMY THALLER

New York, NY 10019

EDUCATION

LUDWIG MAXIMILIAN 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

EXPERIENCE

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 – Present]*

- (In progress) Predicted user-level LTV classification from 30d of user behavior using Vertex AI; assigned the LTV classes via k-means clustering and PCA on long-term user data
- 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
- Analyzed sponsored influencer effectiveness and changed behavior using speech to text APIs and NLP techniques

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
- Presented highly-technical weekly research insights to material scientist colleagues to facilitate research

INDEPENDENT PROJECTS

SPOTIFY ETL AND RECOMMENDATION ALGORITHM

- Leveraged PySpark, scalar-aggregate-reduction optimized SQL queries, & the Spotify Web API to investigate song trends as well as song/genre characteristics via dimensionality-reduction and cosine distance
- Trained a recommendation algorithm using song embeddings trained via Gensim's Word2Vec on 1M Spotify playlists
- Setup an airflow DAG to extract my daily listening history the via Spotify Web API and write it into a personal Postgres Database in a docker container. Then, leveraged the recommendation model to push an AI-recommended playlist to my personal Spotify account each weekq

SKILLS AND TOOLS

Programming Languages (Years of Experience) | Python (5), SQL (2), Java (7), R (1), MATLAB (4), Julia (1)
Python Packages | Pandas, NumPy, Scikit-Learn, Numba, PyTorch, TensorFlow, Keras, PySpark, Regex, WandB, Dask
Data Visualization Software | Data Studio, Looker, Excel/GSheets, Mathematica, Jupyter Notebooks, WandB, Looker, Plotly
Data Engineering Tools | Snowflake, Apache Airflow, Docker, PySpark, DBT