Eric He

eriche98.github.io

✉ eriqqc@gmail.com | ⌨ [EricHe98](#) | 🔗 [Eric-He](#) | ☎ 408-464-9529

Education

New York University	**New York, NY**
M.S. in Data Science (Center for Data Science)	*Sep 2019 - Present*
*B.A. in Mathematics (Courant)	B.S. in Business (Stern)	Minor in Computer Science*	*Aug 2015 - May 2019*

- **Grad GPA:** 4.00/4.00 | **Undergrad GPA:** 3.74/4.00 (Cum Laude)

Professional Experiences

Rocketmiles	**New York, NY**
Senior Data Scientist	*Feb 2021 – Present*

- **Machine Learning:** [search result ranking](#) ([LambdaMART](#)), fraud detection (GBDT), customer lifetime value (BG-NBD), hotel pricing/rewards (proprietary)
- **Live Prediction:** Kubernetes/Docker, Flask, Redis/Couchbase model hosting stack successfully deployed on AWS and on on-premise servers handling millions of requests a day
- **Data Pipelining:** process raw data sources from 3rd parties using dbt and Airflow: Salesforce for customer service, Snowplow for web analytics, Iterable for email marketing
- **Tech:** spearheaded adoption of dbt for data modeling and MLflow for ML experimentation
- **Recruiting:** run assessments and interviews for data science team and intern programs

Rocketmiles	**New York, NY**
Data Scientist	*Oct 2017 – Feb 2021*

- **Ranking search results:** 8% lift in profits with LambdaMART ranking model relative to human-created result scoring equation; data computed with Spark on EMR
- **Fraud detection:** 80% of missed fraud captured using GBDTs with human-superior precision & recall scores; model predictions explained to fraud analysts using SHAP values
- **Inventory pricing:** 10% lift in conversion, 15% lift in gross margin achieved by a profit maximizing inventory pricing formula I derived from a logistic regression conversion model
- **Supply management:** 5% lift in conversion, 21% decrease in error rates from engineering a historical backtester to block error-prone data suppliers; penalty rules and code deployed live

New York University	**New York, NY**
Financial Markets Research Assistant	*May 2017 – Nov 2017*

- [Explored trading strategies](#) using cosine distance/sentiment analysis data derived from the HTML files of corporate annual (10-K) and quarterly (10-Q) statements filed with the SEC
- [Open-sourced cleaned text](#) of 560,000+ statements on the data.world platform with code on Github; the dataset is one of the most popular on the website with over 400 bookmarks

Gyrfalcon Technology	**San Jose, CA**
Deep Learning Intern	*May 2017 – Aug 2017*

- [Replicated Hinton's Model Distillation paper](#) on using "teacher" nets to train "student" nets
- [Implemented teacher-student training](#) with Caffe to test training a VGG-16 CNN as student

Favorite Projects

- **[Learning to Rank Hotels](#)**: benchmark deep learning methods for recommender systems
- **[Markov State Models for Molecular Dynamics](#):** survey paper on probabilistic graphical methods for efficient simulation of the protein folding process

Competencies & Interests

- **Languages:** Python (Torch, XGBoost, Spark), SQL, Bash, R (Tidyverse), Java
- **Stacks:** AWS, Kubernetes | **Databases:** Redshift, Redis, Couchbase
- **Interests:** TV (The Expanse), board games (Arboretum), karaoke, CitiBiking, houseplants