






A passionate, curious-minded and hardworking ML Engineer / Data Scientist. Excited about bringing good ideas from representation learning and sequence modeling over text, images, speech and graphs to production scale.

Currently a second year MS student at NYU's Center for Data Science

(May '23 Exp.)

EXPERIENCE

- 
Spotify
New York, NY
Data Scientist I (Part-Time)
September 2022 - December 2022
 - Delivered usage analytics and case-studies for partners in publishing, licensing and editorial on the effects of inclusion in editorial playlists on global audience growth. Work consisted of open-ended dataset exploration, self-guided analysis and visualization design. A focus on building repeatable analyses resulted in the re-use of created datasets and methods on subsequent projects, several of which are still in use today.
- 
Zillow
New York, NY
Capstone Researcher (Internship)
September 2022 - December 2022
 - Evaluated methods for scalable inductive representation learning over large, sparsely connected multi-modal graph data. Wrote scalable training, evaluation and experiment tracking pipeline to learn neighborhood aggregation functions used for link prediction using Deep Graph Library, Ray Train and PyTorch Lightning, improving recall by 5% over a unimodal baseline.
- 
Spotify
New York, NY
Data Science / Machine Learning Engineering Intern
June 2022 - September 2022
 - Built large-scale entity disambiguation pipeline to match recordings to registered musical works using TensorFlow/Beam/Luigi. Work included performing text vectorization / cosine-similarity measurement on 30M+ daily records and running distributed XGBoost model on a combination of learned and hand-crafted features.
 - Assessed error rates on lyric content across languages and markets. Designed and oversaw annotation trial to evaluate quality control approaches, resulting in a 15% reduction of erroneous lyric assignments on instrumental tracks.
- 
KPMG Data & Analytics
New York, NY
Associate Data Scientist
January 2019 - August 2021
 - Financial Services Client:* Data Engineer on greenfield credit card offering at major NYC financial institution. Worked on small, agile team to rapidly build and scale fault-resistant data processing pipeline capable of handling 10+ million daily transactions using Airflow, PySpark and BigTable.
 - GDPR NLP:* Contributed module to Cognitive Contract Modeling system to utilize RoBERTa embeddings to improve recall by 10% on semantic search over regulatory documentation relating to the EU's General Data Protection Regulation laws.
 - Health Insurance Client:* Built data pipelines to compute cohort statistics for demographic subsets of claims data and provide visualizations for analyst-facing dashboards.
- 
KPMG Cognitive Automation Labs
New York, NY
Summer Intern
June 2018 - August 2018
 - Fine-tuned passage-level topic classification network on privacy documentation using features from pretrained eLMO Bi-LSTM. Achieved F1 score of 0.62 on a held-out set with a selection of ~50 categories on the OPP-115 corpus.

SKILLS

Languages - Python (10 yrs), SQL (6 yrs), Java (2 yrs) Scala (1 yr), C++ (<1 yr)

ML/Data Science - PyTorch, TensorFlow, DeepGraphLibrary, PyTorch-Geometric, NumPy/Pandas,

Data Engineering - Beam/Dataflow, Airflow, Luigi, Spark, Dask

EDUCATION

NYU Center for Data Science 2021 - 2023

M.S. Data Science (Exp.)

NYU Courant Institute of Mathematical Sciences 2014 - 2018

B.A. Computer Science

NYU Leonard Stern School of Business 2014 - 2018

B.S. Statistics

INTERESTS

- Multi-instrumentalist and full-stack musician (writer/performer/producer) _rothie_