David S. Roth

davidroth.me/ ℹ
davidroth6@gmail.com ✉
(201) 615-3785 ☎

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

Currently a second year MS student at NYU's Center for Data Science *(May '23 Exp.)*

EDUCATION

NYU Center for Data Science, *M.S. Data Science (Exp.)*　　　　　　　　　*2021 - 2023*
Machine Learning, Computational Linear Algebra, Inference and Representation, Deep Learning, Reinforcement Learning, Causal Inference, Natural Language Understanding

NYU Courant Institute of Mathematical Sciences, *B.A. Computer Science*　　　*2014 - 2018*
Discrete Math, Linear Algebra, Computer Systems Organization, Algorithms & Data Structures, Large Scale Web Applications, Computer Vision, High Performance Machine Learning

NYU Leonard Stern School of Business, *B.S. Business / Statistics*　　　　　*2014 - 2018*
Probability & Statistics, Linear Regression and Forecasting Models, Regression and Multivariate Data Analysis, Stochastic Processes, Theory of Probability

EXPERIENCE

⌂ Zillow　　　　　　　　　　　　　　　　　　　　　　　　　　　New York, NY
Capstone Researcher (Internship)　　　　　　　　　*September 2022 - December 2022*
- Evaluated methods for scalable inference on graph data. Wrote training, evaluation and inference pipelines to learn dense node representations for node classification and link prediction using gated neighborhood aggregation.

◉ Spotify　　　　　　　　　　　　　　　　　　　　　　　　　　New York, NY
Data Science / Machine Learning Engineering Intern　　　　　*June 2022 - September 2022*
- Built training and inference pipeline for entity disambiguation model to match recordings to registered musical works. Implemented batch inference model with Luigi and Tensorflow using sparse text vectors to score potential match pairs.
- 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 to handle 20+ million daily transaction records using Airflow and Spark.
- *GDPR NLP:* Contributed module to Cognitive Contract Modeling library for using FinBERT text embeddings to do semantic search over regulatory documentation. Improved recall over previous model by 10%.
- *Health Insurance Client:* Built data pipelines to compute cohort statistics for demographic subsets of claims data.

KPMG Cognitive Automation Labs　　　　　　　　　　　　　New York, NY
Summer Intern　　　　　　　　　　　　　　　　　*June 2018 - August 2018*
- Fine-tuned passage-level topic classification network using privacy documentation and features from pretrained eLMo Bi-LSTM. Achieved F1 score of 0.62 on a held-out set with a selection of $\sim$50 categories in the OPP-115 corpus.

SKILLS

- Python (10 yrs), C++ (2 yrs), Scala (1 yr)
- PyTorch, TensorFlow, CUDA/NCCL, Ray, Deep Graph Library, huggingface
- SQL, Beam/Dataflow, Spark, Airflow, Luigi

INTERESTS

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