

David S. Roth

(201) 615-3785 
davidroth6@gmail.com 
github.com/davidsroth/ 
linkedin.com/in/david-s-roth/ 

A passionate, curious-minded and hardworking Data Scientist / ML Engineer. Likes solving big, messy problems with big, messy data. Excited about bringing good ideas from representation learning over text, images and graphs to production scale.






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

(May '23 Exp.)

SKILLS

- **Languages** - Python (9 yrs), SQL (6 yrs), Javascript (2 yrs), Scala (1 yr)
- **ML/Data Science** - PyTorch, TensorFlow, NumPy, Pandas, DGL, huggingface
- **Data Engineering** - Beam, Dataflow, Airflow, Luigi, Spark
- **Cloud** - experience with declarative infra and container-based CI/CD on AWS and GCP.

EXPERIENCE

-  **Spotify- PubSol / DODI** New York, NY
Data Scientist I (Part-Time) September 2022 - Present
 - Delivered usage analytics and case-studies for partners in publishing, licensing and editorial.
-  **Zillow** New York, NY
Capstone Researcher (Internship) September 2022 - Present
 - Evaluated methods for scalable inductive representation learning over large, sparsely connected multi-modal graph data.
-  **Spotify- PubSol / Content Intelligence (3 mos.)** New York, NY
Data Science Intern June 2022 - August 2022
 - Assessed error rates on lyric metadata across languages and markets. Designed and oversaw annotation trial to evaluate quality control approaches.
 - *Embed with Content Intelligence*. Wrote featurization code to perform text vectorization and measure cosine-similarity on 30M+ daily records using PyBeam/Luigi.
-  **KPMG Data & Analytics (2 years, 6 mos.)** New York, NY
Associate Data Scientist January 2019 - August 2021
 - *Health Insurance Client*: Wrote data pipeline to compute cohort statistics for diverse demographic subsets of claims data and provide visualizations for analyst-facing dashboards.
 - *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 5+ million daily transactions using Airflow, PySpark and BigTable.
 - *GDPR NLP*: Leveraged RoBERTa embeddings to improve semantic search over regulatory documentation related to the EU's General Data Protection Regulation laws.
-  **KPMG Cognitive Automation Labs (3 mos.)** New York, NY
Summer Intern June 2018 - August 2018
 - Built feature generation pipeline to perform cleaning, tokenizing and part-of-speech tagging over text data.

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

PROJECTS

- *Remembrall*: An unconventional reminder app using OpenAI's CLIP.

INTERESTS

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