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. 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.

SKILLS

- **Languages** - python (8 yrs), SQL (5 yrs), javascript (2 yrs), Scala (1 yr)
- **ML/Data Science** - NumPy, Pandas, Tensorflow, PyTorch
- **Data Engineering** - Beam, Dataflow, Airflow, Spark, Kafka
- **Cloud** - Experience with container-based development on AWS, GCS and Azure.

EXPERIENCE

Spotify- Publishing Solutions   New York, NY
Data Scientist   *September 2022 - Present*
- Usage analytics and reporting for publishers, licensors and other content partners.

Spotify- Publishing Solutions / Content Intelligence *(3 mos.)*   New York, NY
Data Science Intern   *June 2022 - August 2022*
- Performed analysis of lyric error rates across markets and oversaw annotation process for evaluating approaches to quality filtering.
- Built featurization pipeline in PyBeam to perform text vectorization and calculate cosine similarity for predicting links between recordings and registered compositions. Currently used in production to evaluate ∼50M daily composition/recording pairs.

KPMG Data & Analytics *(2 years, 6 mos.)*   New York, NY
Associate Data Scientist   *January 2019 - August 2021*
- *Financial Services Client:* Data Engineer on greenfield credit card offering at major NY financial institution. Worked on small, agile team to architect and build fault-resistant data processing pipeline capable of handling 1+ million daily transactions using Airflow, PySpark and BigQuery.
- *Telecommunications Client:* Built predictive model to forecast resource requirements on technology initiatives. Used hierarchical time series modeling to improve mean error of timeline and cost estimates by ∼35%.
- *Health Insurance Client:* Built automated data pipeline to accumulate data from diverse groups within the company, calculate derived statistics and display summary data in dashboards for consumption by analysts and data scientists.

KPMG Cognitive Automation Labs *(3 mos.)*   New York, NY
Summer Intern   *June 2018 - August 2018*
- Research/Development Intern on projects expanding core KPMG ML/NLP competencies.
- Built feature generation pipeline for text data, including cleaning, tokenizing, part-of-speech tagging and embedding matching.

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/engineer) ⬤_rothie_