Oliver Spohngellert

ospohngellert@protonmail.com | 917-750-1162
github.com/spohngellert-o | oliverspohngellert.com | linkedin.com/in/spohngellert-o

WORK EXPERIENCE

PEACOCK | Senior Data Scientist
Nov 2022 – Present | New York, NY

- Researched & productionalized content clustering using 3 years of Peacock viewing data (> 10 TB). Clusters used throughout the organization for all marketing, targeting, recommendation efforts (GCP, BigQuery, scikit-learn, Airflow).
- Developed methodology for assigning users to content clusters, accounting for cluster sizes and types (GCP, matplotlib).
- Architected system for inserting new titles into existing clusters on an ongoing basis (BigQuery, Airflow).
- Produced propensity matching pipeline, correcting for systemic non-randomized assignment in Promo tests on Peacock Platform, and producing accurate estimates of Promo on KPIs (CausalML, Airflow, Kubernetes).
- Devised directional content affinity metrics based on Market Basket analysis, used to drive high level strategic decisions.
- Onboarded, mentored, and managed an intern; taught best engineering practices and managed tasks.
- Managed relationships with key stakeholders, presented data for a range of technical expertise, updated team processes.

MAJOR LEAGUE BASEBALL | Machine Learning Engineer
May 2021 – Oct 2022 | New York, NY

- Built end to end classifier for live official scoring of MLB plays in game at 95% accuracy (scikit-learn, seldon, BigQuery).
- Deployed a personalized video recommendation system API for top hitting, pitching, and fielding plays. Used Logistic Regression and Feed Forward Neural Networks to build ranking system (GCP, scikit-learn, kubernetes, argo workflows).
- Developed algorithms for comparing player batting stances with user stances submitted via image (MediaPipe, FastAPI).
- Researched zip code based team affinity using historical article view data (Jupyter, BigQuery).
- Implemented time series classification on skeletal data to determine whether a batter attempted a checked swing.

DEEP HEALTH | Machine Learning Intern
May 2020 – Aug 2020 | Cambridge, MA

- Designed an error analysis pipeline, enhancing understanding of Cancer Detection model (pandas, numpy, matplotlib).
- Developed a system for aggregating cancer annotations, comparing results, generating follow-ups (python, javascript).

NORTHEASTERN NDS2 LAB | Research Assistant
Jan 2019 – Feb 2020 | Boston, MA

- Implemented Recurrent Neural Networks (RNNs) to classify users of IoT systems using exclusively network traffic.
- Engineered features and implemented unsupervised classification models to detect self-propagating malware.

EDUCATION

MASTER OF SCIENCE IN DATA SCIENCE Northeastern University
May 2021 | Boston, MA
GPA: 3.8/4.0 | **Relevant Courses:** Data Science Capstone, Deep Learning, Reinforcement Learning, Supervised Machine Learning, Unsupervised Machine Learning, Statistical Methods in Computing, Data Management, Algorithms

BACHELOR OF SCIENCE IN COMPUTER SCIENCE, MINOR IN MATH Northeastern University
May 2019 | Boston, MA
GPA: 3.73/4.0 | **Relevant Courses:** Software Development, Linear Algebra, Differential Equations, Probability and Statistics

SKILLS

Programming: python, pandas, numpy, sklearn, pytorch, matplotlib, git, GitHub, SQL, linux/unix, bash
Cloud: Google Cloud Platform, BigQuery, VertexAI, Airflow, Google Cloud Storage, Composer, Docker

INTERESTS

Competitive Chess (1748 USCF), Urbanism, Cooking/Baking, Environmentalism