Daniel Ryan Furman

(310) 993 9757 • daniel_furman@berkeley.edu • github.com/daniel-furman

SKILLS

Programming: Proficient in Python (e.g., Numpy, Pandas, Scikit-Learn, Pytorch, Flask), R, Matlab
Version Control/Databases: GIT, GitHub, SQLite, PostgreSQL, Anaconda, Docker
Writing: Published 4 papers in scientific venues, multiple design docs, & code documentation
Cloud: Growing knowledge of GCP & AWS tools for data science and software development

EXPERIENCE

Nov 2021 - Present **Understory.AI - *Data Scientist*** Mountain View, CA

- **Machine Learning Intern**
  - Supported the Software Development team by building and deploying computer vision segmentation models for next-gen land management applications with remote sensing drone imagery. [Link]
  - Delivered an MVP in 6 months that is reproducible to new clients/use-cases, developed demo by deploying 4 segmentation models for inference over 10+ scenes (>80% balanced accuracy).
  - Validated & managed 100+ GB imagery in GCP, developed a hierarchical label search module for 1:1 label to pixel mapping, collaborated on scalable & accurate instance segmentation models (patch-based logistic regressors, U-Nets), applied texture synthesis packages to generate 1000s of synthetic training patches, wrote communicable documentation & structured software with agile practices.

Summer 2019 **Harvey Mudd College - *Data Science Researcher*** Claremont, CA

- Won an NSF grant to design communicable modeling visualizations of climate impacts forecasted by machine learning classifiers of California species distributions (19 feats., ~7k obs.). [DataViz]
- Built and open-sourced a module [PySDMs] for AutoML geo-classification fit via block cross-validation, multi-seed/sample blending, and model search (Catboost, Logistic Regression, etc.), with AUCs>0.95.

Mar 2019 - Present **Contract/Freelance - *Data Science Consultant*** Remote (USA)

- Strategized re-structuring recommendations with **De Castro, West Inc.** for a real estate partnership shifting 11 unevenly shared assets into 3 majority ownerships by building a search algorithm to minimize value & debt stake change, optimal state exhibited <2% change against ~$159mm. [Blog]
- Extracted advanced analytics insights for the **Official World Golf Rankings Ltd.** from over two decades of professional golf tournament data across 21 tours by developing PCA biplots and refining written deliverables, convincing stake-holders to ship updates in Q3 ’22 to boost the system’s fairness.

EDUCATION

Anticipated May 2023 **University of California, Berkeley**, School of Information — Berkeley, CA
Master of Information Management and Systems
Specialization: Machine Learning | Graduate Certificate in Applied Data Science
Teaching: TA for Data Mining @ Berkeley Haas (MBA247.11, Fall 2021) [Syllabus]

Graduated May 2020 **University of Pennsylvania**, College of Arts and Sciences — Philadelphia, PA
Bachelor of Arts with Distinction in Earth Science
Minor: Mathematics • **Cumulative GPA:** 3.79/4.0
Honors: Rose Research Award [Paper], NSF Award #1757952 [Poster], CURF Grant [Write-Up]
Activities: NCAA D1 Golf Team • Experimental Geophysics Lab

PROJECTS

Multi-Platform Mixing for Social Media NLP - Python (NLTK, Hugging Face). [ACL22, In Review]

- Developed a data-centric domain adoption method that improves the portability of NLP models by mixing platforms at training (Instagram & Twitter captions), which significantly boosted F1 performance by 17+% on average for celebrity gender profiling & authorship attribution tasks (SVMs and LRs) under domain shift (Facebook caption test-set).
- Exploration of over 43,000 celebrity Twitter, Instagram & Facebook post linguistics (Rshiny). [DataViz]

SurfCrowds.ai - Python (NumPy, Pandas, Scikit-learn), AWS (Lambda, S3), Dash, Heroku. [Web-App]

- Built a crowd size predictor for surfers in Malibu, CA with Random Forest & Gradient Boosting ensemble regressors deployed as a serverless AWS Lambda function. Embedded models within a Dash web-application that enables my hometown surf community to explore new types of informative AI forecasts.

Kaggle BirdCLEF21 Competition - Python (PyTorch, Pandas, Scikit-Learn), Neptune.ai. [Blog]

- Won solo bronze out of 816 teams for Sound Event Detection with blends of multi-label, split-attention audio RNNs & CatBoost metadata classifiers with non-competition geospatial climate features (397 species classes, ~63k obs.), overcoming a large train to test domain shift & weak labeling (F1=0.61).