

Daniel Ryan Furman

1935 Jefferson Street • San Francisco, CA 94123 • daniel_furman@berkeley.edu

[Portfolio](#) • [LinkedIn](#) • [Google Scholar](#) • [GitHub](#)

EDUCATION

University of California, Berkeley, School of Information – Berkeley, CA

Anticipated May 2023

Master of Information Management and Systems

Specialization: Applied Data Science and Machine Learning | Graduate Certificate in Applied Data Science

Teaching: Lead TA for “Descriptive & Predictive Data Mining”, Berkeley Haas School of Business (Fall 2021) [[Syllabus](#)]

University of Pennsylvania, College of Arts and Sciences – Philadelphia, PA

Graduated May 2020

Bachelor of Arts with Distinction in Earth Science

Minor: Mathematics • **Cumulative GPA:** 3.79/4.0

Honors: Rose Senior Research Award [[Paper](#)], NSF Award #1757952 [[Poster](#)], CURF Climate Action Grant [[Write-Up](#)]

Activities: NCAA D1 Golf Team • Experimental Geophysics Lab

Harvard-Westlake School – Los Angeles, CA

Graduated June 2016

Cumulative GPA: 4.2/4.0 (Weighted) • **ACT:** 34 (Composite Score) • **Activities:** Varsity Golf Team Co-Captain

EXPERIENCE

Engineering Intern, Comon Solutions – Mountain View, CA

November 2021 - Present

- Developed automated image processing and CV modeling pipelines in Python via AWS, refining software products that shape real-time natural resource strategy for land managers with big geospatial data (~ 100's GB files). [[Link](#)]

Data Science Consultant, Contract & Freelance

March 2019 - Present

- Strategized recommendations with **De Castro, West Inc.** for a real estate client restructuring 11 unevenly shared assets into 3 majority ownerships, building a search algorithm that minimized value and debt stake change between partners. Identified an optimal state with under a 2% change against ~\$159mm in total assets. [[Blog](#)]
- Extracted insights for **Official World Golf Rankings Ltd.** from two decades of data (16 feats., 3400 tournaments, 21 tours), developing PCA biplots and refining written project deliverables. These insights convinced OWGR stakeholders to ship updates to the ranking system in Q3 '22 to more accurately capture relative athletic performance.

Data Science Researcher, Harvey Mudd College – Claremont, CA

May - August 2019

- Won an NSF grant to conduct research in mathematical biology, constructing GIFs of species range shifts from ML classifiers trained with bioclimate features & species presence/absence labels (19 feats., ~7k obs.). [[DataViz](#)]
- Built a Python class with CatBoost geo-classification fit via an experimental design with multi-seed/multi-sample frameworks & block cross-validation (AUCs>0.95, domain SOA). Wrote R functions for data ETL & pre-processing.

PROJECTS

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

- Built a surf crowd size predictor for First Point, Malibu, CA with Random Forest & Gradient Boosting ensemble regression deployed as a serverless AWS Lambda function, equipping my surf community with AI-driven forecasts.

Instagram Caption Style Clustering – Python (NLTK, SpaCy), R (shiny, tidyverse). [[DataViz](#)]

- Explored style variability on Instagram posts scraped from three health food influencers (with related topical content), using KMeans clustering on BERT embeddings (stylistic context signals) to predict authorship (NMI = 0.6).

Kaggle BirdCLEF21 Competition – Python (PyTorch, TensorFlow, Keras), Neptune.ai. [[Blog](#)]

- Won solo bronze out of 816 teams for Sound Event Detection with multi-label, split-attention RNNs (ResNest18) and metadata classifiers (397 classes, ~63k obs.), overcoming a large train/test shift and weak labeling (F1=0.61).

Multi-Platform Mixing for Social Media NLP – Python (NLTK, HuggingFace), GCP. [NAACL22, Confidential]

- Developed a data-centric domain adoption method that improves the portability of social media NLP models by mixing multiple platforms at training (Instagram & Twitter), which boosted test-set accuracy by ~50% for celebrity gender profiling on Facebook posts (mimicking domain shift). Scraped a total of ~40k public, verified posts. Stack: Google Natural Language AutoML, BERT-Base, SVMs (tf-idf), Neptune.ai logging, Git for data storage & versioning.