# Daniel R. 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

**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:** CURF Rose Research Award, NSF Award #1757952, CURF Climate Action Grant, EES Hayden Scholar
**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

## RELEVANT EXPERIENCE

**Data Science & Analytics Consultant – Contract**      **March 2019 - Present**
- Extracted product insights for the **Official World Golf Rankings Ltd.** with two decades of proprietary data (16 feats. x 3400 tournaments). Developed analytics such as PCA biplots that drove stake-holders to ship ranking system updates in Q3 2022 that more accurately capture relative athletic performance and strength of field.
- Strategized recommendations with **De Castro, West Inc.** for a real estate partnership restructuring 11 unevenly shared assets into 3 majority ownerships, using a Markov model to minimize stake change between partners (asset value + debt). The best solution incurred ~$2.9mm total stake change against ~$159mm in assets. [Blog]

**Data Mining Teaching Assistant – Berkeley Haas School of Business**      **September 2021 - Present**
- Collaborated on "Descriptive & Predictive Data Mining" for Executive MBAs with Prof. Luyi Yang by helping source data sets for assignments and teaching R programming to students during office hours and lecture. [Syllabus]

**Data Science REU – Harvey Mudd College**      **May 2019 – August 2019 (PT – January 2021**)
- Constructed GIFs visualizing climate change impacts for four marquee CA species from bioclimate features and species presence records (19 feats. x 6800 obs.), using thousands of Global Climate Model rasters. [DataViz]
- Developed a Python class with Catboost & XGBoost ensemble geo-classification fit across an innovative multi-seed/multi-sample pipeline (AUCs>0.95, domain S-O-A). Built R functions for data mining and FSS model selection.

## PROJECTS & INTERESTS

**Instagram Style Clustering with NLP** – Python (NLTK, HuggingFace). [DataViz]
- Explored author attribution on Instagram with captions from health food influencers with highly related topical content, using KMeans on BERT embeddings (capturing stylistic context) to cluster three profiles (NMI = 0.6).

**SurfCrowds.ai** – Python, AWS (Lambda, S3), Dash, Heroku. [Web-App]
- Built a crowd size predictor for Malibu, CA with Random Forest & Gradient Boosting ensemble regression running as an AWS Lambda function, equipping my hometown surf community with new types of AI-driven forecasts.

**Kaggle Competitions** – Python (PyTorch, TF/Keras) , Neptune.ai. [BirdCLEF21 Blog]
- BirdCLEF21: Won solo bronze out of 816 teams for Sound Event Detection with multi-label CNNs (ResNest18) and metadata classifiers (397 classes, ~63,000 obs.), overcoming a large train/test shift and weak labeling (F1=0.61).
- SETI Search: Placed top 17% out of 768 teams for radio anomaly detection by bagging ViTs and CNNs (AUC=0.77).

**Bookshelf** – Range: Why Generalists Triumph in a Specialized World, Superforecasting: The Art and Science of Prediction, Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are.