Yuan **Meng**

(+1) 626-675-4914 | yuan_meng@berkeley.edu | yuan-meng.com | Yuan-Meng | yuanmengds

Technical Skills

Programming Python, SQL, R, Apache Spark, HTML, CSS, JavaScript, Git, Bash, LaTeX, R Markdown

Frameworks/Tools Scikit-Learn, Pandas, NumPy, TensorFlow, Keras, SciPy, PySpark, NLTK, spaCy, PyMC3, Bokeh, Plotly, Flask, Heroku

Stats & Experiments A/B testing, experimental design, causal inference, Bayesian statistics, generalized linear mixed models (GLMMs), survey

Machine Learning Predictive modeling (LightGBM, XGBoost, random forests, neural networks), dimensionality reduction (PCA, t-SNE, UMAP), clustering (K-means, DBSCAN), natural language processing (NLP), recommendation system, time series, survival analysis

Education

University of California, Berkeley | Ph.D. in Cognitive Science (GPA: 3.99/4.0) *Berkeley, CA*

Berkeley Fellowship for Graduate Study (Top 4% Ph.D. students, Berkeley's most prestigious fellowship) *Aug. 2016 - 2022 (EXPECTED)*

- Research areas: Computational models of self-supervised causal learning in children; perpetuation of racial disparities via social learning

Beijing Normal University | B.Sc. in Psychology (GPA: 91.2/100, Rank: 5/103) *Beijing, China*

National Scholarship (Top 1, highest honor for Chinese undergrads), Outstanding Honors Thesis (Top 3%) *Sept. 2012 - Jun. 2016*

Experience

Berkeley Early Learning Lab, UC Berkeley *Berkeley, CA*

Graduate Student Researcher (Committee: Fei Xu, Alison Gopnik, Neil Bramley, Mahesh Srinivasan) *Aug. 2016 - PRESENT*

- Designed and conducted 10+ lab and online (Amazon Mechanical Turk, Prolific, Zoom) behavioral experiments with infants, children, and adults.
- Built Bayesian models of cognition and data analysis in Python (PyMC3) and R (JAGS); taught computational models of cognition (CogSci 131).
- Analyzed (GLMMs, ANOVA, Bayesian statistics, information theory) and visualized (ggplot2, matplotlib, Seaborn) behavioral data using Python and R .
- Trained and managed (manual) 15+ research assistants and 7 summer interns to recruit participants, collect data, and manage lab database.
- Published and presented peer-reviewed research at 6 top cognitive science conferences (including CogSci 2017-2020, CDS 2018, SRCD 2021).

The Data Incubator (2% acceptance rate) *San Francisco, CA*

Data Science Fellow (Mentors: Robert Schroll, Richard Ott) *Jun. - Aug. 2020*

- Built Dash web app (Reopen Navigator) to help users gauge their risk of contracting COVID-19 and having severe outcomes; pitched to stakeholders.
- Scored 1.0/1.0 in 8 advanced ML projects on web scraping, NLP, network analysis, SQL, Spark, TensorFlow, time series, and ETL pipelines.

Thinkful Data Science Program *New York City, NY*

Student (Mentor: Davin Kaing) *Sept. 2019 - Jun. 2020*

- In-depth training in data structures and algorithms, data wrangling and visualization, A/B testing, supervised and unsupervised learning, etc..
- Scored 9.6/10 in 5 capstone projects, including scraping and topic modeling of Nightwish lyrics and building a group movie recommender.

Selected Projects

Reopen Navigator (Dash App) | Python (LightGBM), Dash, Heroku *San Francisco, CA*

Use case: Gauge COVID-19 infection risk and outcome severity based on user social bubble, state, and conditions *Jun. - Aug. 2020*

- Scraped 20.6 million COVID-19 case and viral test data from state reports; reviewed 50+ medical journals to estimate average COVID-19 infectivity.
- Trained and deployed LightGBM classifier (F1 score = .84) to predict new user's treatment outcome based on state, demographics, and conditions.

Movie Night | Python (Surprise, SciPy) *New York City, NY*

Use case: Help a group of users with varying tastes find the best movie for their movie night *May - Jun. 2020*

- Built collaborative filtering recommender for individual users using SVD++ on MovieLens data, outperforming baseline model (RMSE = .87 and 1.4).
- Reviewed 15+ papers on "cold start"; solved it by asking new users to rate 20 movies with highest harmonic means of rating entropy and log frequency.
- Extended recommender system to user group by aggregating individual recommendations ("average-without-misery") or merging user profiles.

Data Mining, Database, & NLP Projects | Spark, SQL, Python (Scikit-Learn) *San Francisco, CA*

Compilation: The Data Incubator projects involving web scraping, distributed computing, and large-scale networks *Jun. - Aug. 2020*

- Web-scraped 100,000 photo captions in New York Social Diaries and applied NLP and network analysis to constructing social graphs of NYC elites.
- Processed 100 GBs of Stack Overflow posts with Apache Spark and used Word2vec in Spark ML to identify top 25 closest synonyms to "ggplot2".
- Devised SQL database for Yelp restaurants; built ETL pipelines to predict ratings from reviews, including n-gram model with Bayesian smoothing.

How Do Disparities Reproduce Themselves? | Python (PyMC3, Scikit-Learn) *Berkeley, CA*

Research: Examine how social disparities may be perpetuated by rational inference and look for mitigation strategies *Sept. 2019 - PRESENT*

- Created game to simulate proactive policing where a knowledgeable, cost-efficient agent searches groups for target trait (e.g., crime) at different rates.
- Compared 3 Bayesian models inferring population hit rate from agent's search rate and sample hit rate (using one or both) with human intuitions.
- Found showing search outcome stopped people from reproducing disparities by updating inferred hit rate; presented at top venues (CogSci20, SRCD21).