Yiwen (Ewen) Dai

ewendai@berkeley.edu | linkedin.com/in/ewendai | ewendai.github.io | Berkeley, CA | (510) 598-7025

Education

University of California, Berkeley
Berkeley, CA

Data Science B.A. and Statistics B.A. [Cumulative GPA: 3.522/4.00] — Fall 2017 - Fall 2020

- Regents' and Chancellor's Scholarship (Top 2% of Accepted UC Berkeley Students)
- Technical Coursework: Numerical Analysis, Probability, Concepts of Statistics, Linear Modelling, Stochastic Processes, Data Structures, Efficient Algorithms & Intractable Problems, Artificial Intelligence, Principles & Techniques of Data Science, Optimization

Selected Projects

Taxi Ride Durations (DS 100) — Spring 2019

- Created a regression model that predicts the travel time of a taxi ride in New York using yellow taxi trip records in January 2016 as published by the NYC Taxi and Limousine Commission (Python)
- Utilized feature engineering, cross validation, and regularization to determine the best model from linear and decision tree regressions

Pacman AI (CS 188) — Spring 2019

- Implement search and multiagent algorithms along with reinforcement learning, probabilistic inference, and basic machine learning techniques to help Pacman win the game (Python)

Selected Activities

Data Science Consultant
Berkeley, CA

Student Association for Applied Statistics (SAAS) — Fall 2019 - Present

- Partnership with and Data Science Intern at Data Secrets
- Data Cleaning, Data Preprocessing and Vectorization using NLP
- Utilized clustering algorithms (DBSCAN and HDBSCAN) as well as engaging in hyper-parameter tuning to test for best results (Python)

Data Science Consultant
Berkeley, CA

Data Science Society (DSS) — Fall 2019 - Present

- Partnership with and Data Science Intern at Advisary
- Created Schema on DynamoDB for Consultant Profile
- Coded ML Recommendation Engine, along with the necessary pseudo-data creation functions (Python)

Prob 140 Group Tutor
Berkeley, CA

UC Berkeley Statistics Department — Fall 2019 - Present

- Facilitate in teaching Statistics 140 (Probability for Data Science) by answering questions and explaining probability concepts in office hours
- Grade weekly assignments

Undergraduate Researcher
Berkeley, CA

Goodly Labs, Public Editor Project — Spring 2019 - Present

- Lead the Monitoring and System Calibration (MASC) Subteam of 5-9 members and work to ensure high quality user responses
- Create data cleaning and analysis pipeline to analyze variance associated with each credibility indicator, from which article scores are calculated
- Code functions to create pseudo-data for testing purposes (Python)

Software Engineering Intern
San Francisco, CA

Influxdata, Query Team — Summer 2019

- Add strings and regex functionality to Flux query language (GoLang)
- Implement transformation functions from InfluxQL into Flux (GoLang)
- Create new stream type and allow for writing data from one stream to another in pure Flux (GoLang)

Skills

Programming Languages: Python (incl. Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn), GoLang, Java, R (incl. ggplot2, dyplr), Matlab, SQLite, HTML, CSS, Javascript
Foreign Languages: Mandarin Chinese
Work Eligibility: Eligible to work in the U.S. with no restrictions