

Sean Wei

858-705-1272 | seanwei2001@berkeley.edu | LinkedIn: seanwei2001

Education

University of California, Berkeley

M.S. in Data Science

B.A. in Data Science, Domain Emphasis in Cognition

GPA: 3.91/4.00 (Dean's List Fall 2021)

Coursework: CS 189 - Introduction to Machine Learning, Data C100 - Principles and Techniques of Data Science, Data 144 - Data Mining and Analytics, Stat C140 - Probability for Data Science, Data 101 - Data Engineering, Data C102 - Data, Inference, and Decisions, Info 159 - Natural Language Processing, CS 188 - Introduction to AI

Expected Graduation: Dec 2024

Expected Graduation: May 2023

Skills

Tools: Python (NumPy, Pandas, sklearn), MySQL, postgresSQL, noSQL, Java, Redshift, AWS, Github, Jupyter, MongoDB

Skills: Machine Learning (Random Forests, XGBoost, Regression, Neural Networks, Clustering, k-Nearest), Feature Engineering (One-Hot Encoding, PCA), Data Visualization (Tableau, Matplotlib, Seaborn), Data Engineering, Hypothesis Testing, Dashboard

Professional Experience

Data Analyst Intern Experian

May 2022 - Aug 2022

- Generated an XGBoost classifier to predict which credit card promotions would yield the highest conversion rate for over 1,000,000 Experian customers pulled from AWS Redshift, based on their consumer behavior, financial history, and demographic information. Early results indicated a 200% increase in response rate.

Project Manager Data Science Society

Sep 2021 - Present

- Led a group of 5 analysts, in partnership with a major sports agency, in completing a Streamlit dashboard which utilized k-means clustering and random forests to match 1000+ college athletes with the best companies for sponsorship endorsements, based on their school, sport, and social media presence.

Data Science Intern Basil Labs

May 2021 - Aug 2021

- Collaborated with other interns to mine online reviews and news for 2000+ small businesses around the world, organize the data with MongoDB, and use natural language processing and sentiment analysis to generate interactive dashboards, apps, and business intelligence reports for several clients.

Undergraduate Student Instructor UC Berkeley Division of Computing, Data Science, and Society

Jan 2020 - Present

- Led sections of 30+ students through discussions, coding exercises, and office hours towards a deeper understanding of Python, data manipulation, probability, hypothesis testing, machine learning, and other data science techniques.

Vice President of Projects Sports Analytics Group at Berkeley

Sep 2019 - Present

- Established partnerships with 10+ professional and collegiate sports organizations to conduct data science and machine learning consulting projects on game strategy and player development. Led committee-wide meetings and connected with team members to ensure project leads maintained sufficient progress towards deliverables.

Projects

Predictive Smoothing Sea Warden

Spring 2022

- Worked closely with a team to create multiple exponential weighted moving average algorithms and generate unique evaluation metrics to smooth Sea Warden's time series predictions and yield 60% more accurate and consistent results.

Pokemon Data Mining and Analytics Kaggle Competition

Fall 2021

- Worked with a group to train a neural network, logistic regression, and random forest model on 1,032 Pokemon and 50,000 battles to classify and predict 10,000 battle results with a validation accuracy of 94%.

Text Analysis and Voting Rights ReThink Media

Fall 2021

- With a team, utilized Gaussian Naive Bayes, Decision Trees, Random Forests and various ensemble methods to classify 12000+ news articles as news or opinion, and by author gender with 97% test accuracy.

Understanding Low Income Housing to Decarbonize It Natural Resources Defense Council

Spring 2021

- Collaborated with a team to mine and analyze census data in order to classify 540,000+ low-income households in numerous states by income, age, energy cost, family size, and carbon reduction potential.