

Nancy Yinuo Xu
(626) 427-5725 • yinuoxu54@berkeley.edu

EDUCATION

University of California, Berkeley	Berkeley, CA
B.A. in Data Science	Spring 2023
B.A. in Statistics	
University of Michigan	Ann Arbor, MI
PhD in Information	2023 – Present

RESEARCH EXPERIENCE

UC Berkeley School of Information	Berkeley, CA
<i>Research Assistant</i>	Sept. 2022 – May 2023

- Researched and developed statistical tests to detect semantic variation and change using contextual and non-contextual NLP methods.
- Conducted a literature review on the theoretical framework of independence tests, evaluation schemes and data synthesis, and methods of using contextual models to detect semantic changes.

UC Berkeley School of Education	Berkeley, CA
<i>Research Assistant with Professor Zachary Pardos</i>	Sept. 2021 – Present

- Research on improving student course recommendations through mining for course similarity signals in detailed enrollment data.
- Build and automate training and experiments for skip-gram, fasttext, BERT, and log-linear models on Gensim and PyTorch to output course recommendations.
- Create novel analytical explanatory frameworks and visualizations to increase the interpretability of models and course recommendations.
- First author of the resulting papers submitted to AIED and EDM conference.

UC Berkeley Department of Sociology	Berkeley, CA
<i>Research Assistant with Professor Heather Haveman</i>	Sept. 2021 – Present

- Research, implement, and compare various NLP models and data processing techniques to build tools for computational literature review.
- Trained deep learning models (BERT, Longformer) and created ML pipeline on PyTorch to build multi-class classifiers for JSTOR social science articles.
- Design unsupervised methods (word embedding, hierarchical clustering) to develop concept clusters and track shifts in theoretical sociology paradigms over time.

PROFESSIONAL EXPERIENCE

Evisort	San Mateo, CA
<i>Data Science Intern, NLP</i>	May. 2021 – Aug. 2021

- Built an algorithm to redact sensitive information in contracts with Google DLP API, and libraries such as regex, XML, docx, and Python.

- Collaborated closely with software and legal teams to build, train, and deploy Light GBM and bi-LSTM NER Tensorflow models that extract fields and provisions in contracts to a modern SaaS platform. The integrated models are used by clients on the SaaS platform to gain valuable insights on legal contracts.
- Designed PowerBI dashboards to analyze document metadata and search usage patterns to provide clients insights on legal contracts by querying and transforming AWS data with PostgreSQL.

TEACHING EXPERIENCE

UC Berkeley School of Information

Berkeley, CA

Teaching Assistant, Natural Language Processing (Info 159/259)

Jan. 2023 - May 2023

- Create homework assignments and quizzes on NLP topics for 260 undergraduate and graduate students, and teach content and assignment-related questions during office hours.

UC Berkeley Division of Data Science

Berkeley, CA

Undergraduate Student Instructor, Data 100

May 2022 - Dec. 2022

- Teach 30 students in discussion sections, present materials on the principles and techniques of data science, and answer questions during office hours in an upper-division data science course.
- Managed logistics of discussion sections and weekly announcements in a 300-student class.
- Received effectiveness rating 6.2/7 on course evaluations.

UC Berkeley EECS, Rausser College of Natural Resources

Berkeley, CA

Student Facilitator, Technology and Social Impact (EECS 198)

Sept. 2020 - May 2023

- Present materials and facilitate discussions around the ethics of technology in various fields, such as social media, government, art, and education.
- Led a group of 4 student facilitators to develop the course assignments and [syllabus](#), focusing on critical engagement with course content outside the classroom.

PUBLICATIONS

Y. Xu and Z. A. Pardos. [Mining detailed course transaction records for semantic information](#). In Proceedings of the 16th International Conference on Educational Data Mining, pages 388–395, Bengaluru, India, July 2023. International Educational Data Mining Society.

Towards Data Science – “[Deep Learning Reviews the Literature](#)” by Jaren Haber, Thomas Lu, and Nancy Xu, Jan. 18, 2022.

VOLUNTEER SERVICES

Activism Always

Berkeley, CA

Data Engineer Intern

Feb. 2021 — Aug. 2021

- Created an AI web application for nonprofits to visualize social media data.
- Designed database schema, and automated social media data collection and cleaning on MongoDB; built social network graphs for influencer identification on scraped Twitter data and deployed an NLP Sentiment Analysis model with PyTorch and Flair; designed Flask web app to

present recommendations for nonprofits.

DaanMatch

Berkeley, CA

Data Science Intern

Sept. 2020 — May 2021

- Mined and cleaned data to identify key metrics examining the relationship between the prevalence of malaria, healthcare resources, and land use in Northeast India; leveraged Python and Tableau to map vulnerable populations in Northeast India to help nonprofits provide needed donations.
- Deployed and trained a multi-class LDA classifier to classify NGOs' sustainable development goals, and deployed the model to a web app on Flask.

City of Paterson, New Jersey

Berkeley, CA

Data Research Analyst

Sept. 2020 — Dec. 2020

- Analyzed and visualized public sentiment on evictions and its correlation with poverty rates and housing policies in neighboring cities.
- Web scraped Twitter and developed N-gram and sentiment analysis of tweets on evictions in Paterson, New Jersey.
- Established key metrics for analyzing public sentiment on COVID-induced evictions and eviction policies.

LEADERSHIP & OUTREACH

Tech + Social Impact

Berkeley, CA

Organization Lead

Sept. 2019 — May 2023

- Organized semesterly speaker events and communicate with nonprofits for partnerships.
- Coordinated club meetings, and provide resources and guidelines to student project managers to create tech for social good projects.
- Led data science project team building tools for hate speech detection.

Gifts at Berkeley

Berkeley, CA

Co-founder, Director of Outreach

Sept. 2019 — Aug. 2022

- Formed partnerships with nonprofit CoachArt, and managed a team of student volunteers to host arts & crafts lessons for children with special needs.

SKILLS

Programming: Python (3+ years), SQL (3+ years), R (3+ years), Java (1+ years)

Data Visualization: Tableau, PowerBI, Excel, Seaborn, Matplotlib, ggplot2

Libraries: Pandas, Numpy, Scikit-learn, SciPy, PyTorch, Tensorflow, Flair, NLTK, Gensim, Tweepy, R Shiny, dplyr, Flask

Modeling: linear/logistics regressions, decision trees, PCA, k-means clustering, k-nearest neighbors, A/B testing, RNN, BERT, CNN, NER, LDA