

Erica Xia

<https://ericaxia.github.io/>

Email : ericaxia@usc.edu

Mobile : +1-925-535-1045

EDUCATION

- **University of Southern California** Los Angeles, CA
M.S. in Applied Data Science Jan. 2020 – Dec. 2021 (*exp.*)
- **University of California, Davis** Davis, CA
B.A. in Statistics and Economics Sep. 2014 – Mar. 2019

EXPERIENCE

- **Research Assistant, Epicenter for Disease Dynamics** Davis, CA
University of California, Davis Dec. 2018 - Dec. 2019
 - **NLP Relation Extraction Project:** Implemented Snorkel machine learning framework to extract virus-host relations from research literature, including data preprocessing, model training, tuning hyperparameters, and error analysis. Assisted with drafting manuscript.
 - **Topic Modeling Project:** Performed tuning and analysis of Latent Dirichlet Allocation model and topic extraction. Created bipartite network visualizations. Advised on manuscript revision.
 - **Data Wrangling:** Conducted data cleaning, standardization, and organization of species taxonomy datasets. Scraped and queried data from web.
- **Accounting and Finance Intern** San Francisco, CA
SocialChorus (SaaS Company) June 2017 - Aug. 2017
 - **Accounting Functions:** Managed A/R and A/P accounting, billing, and collections. Assisted in month-end close of the P&L and GL. Completed bank reconciliations and financial reports.
 - **Data Organization:** Managed customer data using Salesforce. Collaborated with finance team on data cleaning and visualization projects.
 - **Concur Implementation:** Handled company-wide transition to Concur expense management software, wrote guides on company website, and led instructional workshops.
- **Financial Training Program Intern** Sacramento, CA
Morgan Stanley Apr. 2017 - June 2017
 - **Market Research and Presentations:** Researched and presented stock pitches. Wrote sector analysis reports. Trained in stock analysis and portfolio construction.

PROJECTS

- **Edtera Student-Teacher App - USC Directed Research Project (2020):** Collaborate on back-end RESTful web development. Implement GraphQL server. Research and plan ML features to implement.
- **Tracking Health and Nutrition Signals - USC CKIDS Project (2020):** Web scrape Instagram social media data. Assist with pertinent NLP, predictive modeling, and image analysis.
- **Coding Dojo Data Science Hackathon, First Place (2019):** Performed exploratory analysis and feature engineering. Built ML classifier predicting college student enrollment to 93% accuracy.

SKILLS

- Python: NumPy, Pandas, Sci-kit Learn (sklearn), Matplotlib, Flask, PyTorch
- R: Tidyverse (ggplot2, dplyr, tidyr, purrr, forcats)
- Statistical modeling, data analytics, machine learning (KNN, regression, random forest, RNN, LSTM)
- Relational databases: SQL, GraphQL, MongoDB (NoSQL)
- UNIX, Bash scripts, Ubuntu Linux, Git, Docker

PUBLICATIONS (IN PROGRESS)

- **Pandit P., Xia E., Johnson C.K.** Global Perspective on One Health: Varying Priorities and Health Issues.