

# Erica Xia

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

Email : [ericaxia@usc.edu](mailto:ericaxia@usc.edu)

Mobile : +1-925-535-1045

## EDUCATION

---

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

## EXPERIENCE

---

- **UC Davis (Epicenter for Disease Dynamics)** Davis, CA  
*Research Assistant* Dec. 2018 - Sep. 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:** Assisted with tuning and analysis of Latent Dirichlet Allocation model.
  - **Data Wrangling:** Conducted data cleaning and organization of species taxonomy datasets. Scraped and queried data from web.
- **SocialChorus (SaaS Company)** San Francisco, CA  
*Accounting Intern* 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 projects including data cleaning and visualization.
  - **Concur Implementation:** Assisted with company-wide transition to Concur expense management software, wrote guides on company website, and led instructional workshops.
- **Morgan Stanley (Training Program)** Sacramento, CA  
*Financial Training Program Intern* Apr. 2017 - June 2017
  - **Seminar Training:** Attended seminars to train in stock analysis and portfolio construction.
  - **Presentations:** Collaborated with team to present stock pitches and write sector analysis report.

## SKILLS

---

- Python: NumPy, Pandas, Matplotlib, Sci-kit Learn (sklearn), Tensorflow, PyTorch, SpaCy
- SQL, relational databases, Postgres
- Unix commands and Bash scripting in Linux
- Web scraping, regular expressions, DOM/CSS/HTML, parsing JSON, and information extraction
- Statistical modeling in Python and R

## PROJECTS

---

- **Coding Dojo Hackathon, First Place (2019):** Built ML classifier predicting student enrollment to 93% accuracy, in addition to exploring and engineering features.
- **Scikit-learn Open Source Sprint (2019):** Contributed to documentation and minor issue fixes.

## RELEVANT COURSEWORK

---

- **USC:** Data Science at Scale, Foundations of Data Management
- **UC Davis:** Mathematical Statistics, Linear Algebra, Calculus, Statistical Data Science, Econometrics, Analysis of Categorical Data, Regression Analysis, Analysis of Variance, Multivariate Statistics