

Irene Yang

xyang68@dons.usfca.edu | 415-624-4393 | San Francisco, CA | LinkedIn | Github | Portfolio

Education

MS in Data Science - University of San Francisco

July 2018 – July 2019

- **Courses:** Machine Learning, Natural Language Processing, Relational Databases (SQL), Experimental Design (AB Testing), Deep Learning, Linear Regression and Distributed Computing

BA in Economics - Xiamen University

Sep. 2014 – June 2018

- **Courses:** Statistics, Econometrics, Linear Algebra, Probability and Calculus

Experience

Data Scientist Intern | Reputation.com | Redwood City, CA

Oct. 2018 – Present

"A technology pioneer for online reputation management and customer experience management." Source: WIKIPEDIA

- Used ML & NLP techniques (Python, PyTorch, SQL) to assist company in managing online reputation.
- Improved the sentiment classification accuracy from 88% to 93% with a deep learning model using LSTM and self-attention. Developed a sentiment visualization for trigger words to interpret model results.
- Constructed a multi-gram keyword extraction tool to identify 100 customer concerns from reviews. Improved the accuracy of previous tool by 10% through syntactic dependency analysis. (Blog)
- Automated insights report including client's strengths and weaknesses for marketing team. Reduced time needed by 50%.

Data Analyst Intern | Vanke | Fujian, China

Mar. 2018 – June 2018

"A Fortune Global 500 company with \$44 Billion market cap." Source: WIKIPEDIA

- Conducted data acquisition & analysis (Python, SQL, Tableau) to support real-estate investment decision making.
- Built a web scraping tool to collect companies' public geographical and financial data. Improved data collection efficiency.
- Analyzed and visualized local company geographical and financial patterns to support office building design and investment.

Research Assistant in Econometrics | Xiamen University | Fujian, China

Oct. 2017 – June 2018

- Analyzed social media effect on Initial Coin Offering (ICO) funding using regression for over 1000 companies.
- Implemented causal inference with tree-based and LASSO-based machine learning methods on high dimensional data.

Projects

Distributed NYC Parking Tickets Clustering Analysis (Paper Accepted) [Spark, MongoDB, AWS]

- Clustered 8G parking tickets based on vehicle characteristics and travel time using Spark on AWS EMR. Compared the cost between different EMR configurations. Reduced the time of data preprocessing and model training by 60%. (Github)

Mobile In-App Purchase Prediction [Python, GCP]

- Predicted user purchase within next 7 days with recall 0.88 using over 20 GB user session data and a stack of tree-based models (Top 5 Team). (Presentation)

Newsfeed Product Development [Python, AWS]

- An analytic newsfeed product with topic controversy and author impact analysis. Techniques including topic modeling, sentiment analysis and model deployment. (Github)

Programming Skills

Languages: Python, PyTorch, SQL(PostgreSQL, Redshift), NoSQL(MongoDB), R

Big Data Techniques & Tools: AWS(S3, EC2, EMR), GCP, Git, Spark, Hive, Tableau

Statistics: Experimental Design (AB Testing), Hypothesis Testing, Time Series Analysis, Regression

Machine Learning: NLP, Random Forest, Gradient Boosting, Clustering, Neural Network