

# Irene Yang

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

## 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: WIKIEPDIA

- Used ML & NLP techniques (Python, PyTorch, SQL) to assist compapny in managing online reputation.
- **Sentiment Analysis:** Classified review sentiment using Random Forest and XGBoost. Improved the accuracy to 93% with a deep learning sentiment classifier using LSTM and self-attention.
- **Natural Language Processing Tooling (Blog):** Constructed a multi-gram keyword extraction tool to identify 100 customer concerns from reviews. Improved previous tool accuracy by 10% through syntactic dependency analysis.
- **Operational Insights Analysis:** Automated insights report including client's strengths and weaknesses for product and marketing team by optimizing SQL query, reduced time needed by 50%.
- **Anomaly Detection:** Built a review volume monitor using statistical testing to detect anomaly increases and drops of review volume. Supported data quality assurance.

### Data Analyst Intern | Vanke | Fujian, China

Mar. 2018 – June 2018

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

- Use data collection & analysis (Python, SQL, Tableau) to support real-estate investment decision making.
- **Web Scraping:** Developed web scraping tool to collect prospective customers data including public geographical and financial data. Improved data collection efficiency.
- **Analytics & Data Visualization:** Analyzed customers features to support office building design. Identified and visualized office building geographical trends.

### Research Assistant in Econometrics | Xiamen University | Fujian, China

Oct. 2017 – June 2018

- Applied Econometric techniques to analyze the social media effect on ICO funding; Implemented ML causal inference method.
- **Data Collection & Regression:** Analyzed Initial Coin Offering (ICO) projects funding success for over 1000 companies with 12 features from Twitter. The Project included data acquisition, visualization and modeling.
- **Causal Inference:** Implemented causal inference with tree-based and LASSO-based machine learning methods on high dimensional data. Identified and visualized heterogeneous treatment effect on empirical 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)

### Newsfeed Product Development [Python, AWS]

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

## 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