# Liying Li

**Cell: (+1)415-792-8329 E-mail: lli93@dons.usfca.edu**

## Education

2018.07-2019.06 (Expected) **College of Arts & Science, University of San Francisco** — *Master of Science in Data Science* — San Francisco, CA

- Relevant courses: Machine Learning, Advanced Machine Learning, Problem Solving with Python, Data Acquisition, Relational Databases, Distributed Computing

2014.07-2018.07 **School of Economics & Management, Tsinghua University** — *Bachelor in Economics and Finance* — Beijing, China

- Relevant courses: Econometrics, C++, Applied Time Series Analysis, Applied Statistics and Data Analytics, Stochastic Mathematics Methods

## Internship

2018.10-2019.06 (Expected) ***Beam Solutions*** — *Data Scientist Intern* — San Francisco, CA

**US census application project**

- Assisted in building the US-census service by retrieving data from US-census website using a REST API and Elasticsearch. Created a Docker image running as a container using Gunicorn and AWS EC2 instances to deploy the application.

**Anomaly detection for broker-dealer transaction**

- Engineered new features on broker-dealer transaction dataset andmodeled the dataset with unsupervised clustering algorithms, including K-Means, Mean Shift and Tree Ensemble methods.
- Built Isolation Forest model for detecting the abnormal transactions. Built several models by using different number of features and 2,000 out of 60 thousand observations were indicated as anomalies.
- Analyzed modeling results by checking the characteristics of the anomalies based on possible rules of anomaly transactions, which showed the anomalies could be the transactions for money laundering with high probability.

2017.12-2018.06 ***DiDi*** — *Data Analyst Intern* — Beijing, China

- Worked on strategy solutions to attract inactive users while supporting testing strategy in various cities. Achieved 10% re-activation of inactive users.
- Analyzed relevant app data in SQL and Hive.
- Supported roll out of the re-activation strategy across ten cities to increase. drivers' service hour by 2 hours per day on average.
- Assisted with screening fraudulent behavior using Tableau.

2017.6-2017.10 ***Shenzhen Yan Sheng Asset Management Co., Ltd.*** — *Quantitative Intern* — Beijing, China

- Analyzed historical data for several futures products with a large capital investment such ass reinforced steel bar, polypropylene, and cooking coal.
- Developed and applied futures strategy based on mean reversion of polypropylene for a five-minute line.
- Increased net profit and reduced retracement by adjusting parameter sensitivity resulting in 100% return rate in the historical data test.

## Projects

2019.1-2019.3 **Prediction of AQI based on spark Machine Learning** — San Francisco, CA

- Completed feature engineering and preprocessing on air pollutant data set using spark distributed computing on EMR clusters.
- Used Spark Machine Learning package to model Random Forest Classifier and Logistic Regression Classifier, and achieved accuracy greater than 0.8 of the final model.
- **Tools**: Python, Spark (SparkSQL, SparkML, pyspark), AWS EMR, Scikit-Learn

2018.11 **Canadian Bankruptcy Rate Forecasting** — San Francisco, CA

- Predicted Canadian monthly bankruptcy rates based on multivariate time series models.
- **Tools**: R, Time Series(Box-Jenkins, SARIMAX, Holt-Winters, VARX)

2018.08 **Twitter Sentiment Analysis** — San Francisco, CA

- Analyzed sentiment for the automated-fetched tweets. Generated the results ranked according to the scores and shown as a digested list with different colors.
- **Tools**: Tweepy, Jinja2, Flask, REST API

## Other Information

- Skills: Python, R, SQL, Spark (Spark SQL, SparkML, pyspark), AWS(S3, EC2, EMR, Sagemaker), NoSQL(MongoDB), Git, C++
- Language: English(Fluent), Mandarin(Native), Russian(Basic)