Zhi Li

San Francisco, CA, 94105 | (530)-400-3731 | zli142@usfca.edu

EDUCATION

University of San Francisco | M.S in Data Science **Jul. 2018 – Expected Jun. 2017**

Coursework: Advanced Machine Learning, Data Visualization, Distributed Computing, Data Acquisition, Time Series Analysis, Computational Statistics, A/B Testing, Experiment Design, Time Series Analysis

University of Florida | Graduate Study in Environmental Horticulture **Aug. 2017 – May. 2018**

Coursework: Intro to Computer Programming with R, Introduction to Applied Statistics

University of California, Davis | M.S in Agronomy & Horticulture **Sep. 2015 – Jun. 2017**

Coursework: Applied Multivariate Modeling, Advanced Plant Breeding

Scholarship : Graduate Student Fellowship (2015-2017), Pauleden F. and Dorathea Knowles Scholarship (2016)

Northeast Forestry University | B.S. in Forestry Resources **Sep. 2011 – Jun. 2015**

Exchange Program : National Chung Hsing University(Sep. 2013 – Jan. 2014), Harbin Institute of Technology (2011 – 2012)
Minor in Accounting
Scholarship : Merits Academic Scholarship (2011-2015)

WORK EXPERIENCE

Data Science Intern | Fair.com **Oct. 2018 – Current**

- *Implemented data extraction and feature engineering using Snowflake ETL, SQL and Python*
- *Created vehicle vectorization model & vehicle recommendation engine with EM clustering and word2vec*

PROJECTS

Forecasting Canadian National Bankruptcy Rates | Course Project

- *Predicted Canadian monthly bankruptcy using multivariate time series model*

Tools: R, *Time Series, Box-Jenkins, ARIMAX/ SARIMAX, Holt-Winters, VAR/VARX*

Twitter Sentiment Analysis | Course Project

- *Pull data from Twitter using tweep wrapper, analyzed and visualized the intensity of sentiment of Twitter posts.*
- *Built a web server running on AWS to display the most recent 100 tweets from a given user and the list of users followed by a given user.*

Tools: *Python, AWS(EC2), Twitter API, Jinjia2, vaderSentiment library, tweepy library*

Recommending Articles with Word Embedding | Course Project

- *Calculated the five most similar articles for each article in the BBC dataset, based on the Euclidean distance between each document' centroid.*
- *Built a web server running on AWS to recommend the five most similar articles for each one.*

Tools: *Python, AWS(EC2), Jinjia2, word2vec, tokenization*

PROFESSIONAL SKILLS

- **Programming and Tools:** *Python, R, AWS(EC2, S3, EMR), Bash, Git*
- **Database and Distributed Computing:** *SQL(PostgreSQL), NoSQL(Mongo DB), Spark*
- **Languages:** *English (Bilingual Working Proficiency), Mandarin Chinese (native)*