# ZELONG (ERIC) ZHANG

Santa Clara, CA | (631)605-2687 | er1czz.github.io | linkedin.com/in/zhangzelong | zelongz@lsu.edu

## SKILLS

**Programming**: Proficient in Python, Bash, and High-Performance Computing (HPC, AWS)
Familiar with SQL, Tcl, HTML, and CSS

**ML libraries:** Matplotlib, Bokeh, scikit-learn, NLP, fastai, PyTorch, and TensorFlow

**Quantitative:** Statistics & Probability, Linear Algebra, Multivariable Calculus, and Optimization

## EXPERIENCE

**Fellow,** Insight Data Science, San Francisco, CA — Sep 2020 - Present

- Perform NLP to analyze user data of THE RUN EXPERIENCE ™, a fitness app from Ongo Science
- Predict user subscription using Logistic Regression and Random Forest: achieved $F_1$ score 0.8
- Identify key factors of user conversion such as text length, engagement frequency, sentiment, etc.

**Research Assistant**, Geology & Geophysics, Louisiana State University — Sep 2014 – Aug 2020

- Investigated materials corrosion by collaborative research with experts in glass, metal, and ceramics
- Developed predictive models using time-series data to improve nuclear materials disposal safety
- Produced an award-winning short film showcasing our research synergy (US Dept. of Energy, 2019)
- Investigated oil recovery from shale nanopores using molecular dynamics simulation
- Identified optimal conditions (e.g. temperature, salinity, etc.) for oil production
- Initiated and coordinated collaborations with Shell Netherlands and Citrine Informatics

**Teaching Assistant**, Geology & Geophysics, Louisiana State University — Jan 2020 – May 2020

- Produced lab lecture videos for 15 non-major college students to continue their study remotely
- Re-designed lab courses and built a website on GitHub hosting class materials for remote access

**Research Assistant**, Geosciences, Stony Brook University, NY — Aug 2011 – May 2014

- Developed a methodology using solid-state NMR to characterize organophosphates in calcite
- Processed time-series signals into frequency-based spectra by Fourier-transform

## PROJECTS

**Materials Stability Prediction**, Citrine Informatics, Redwood City, CA — Aug 2020 - Sep 2020

- Applied a stacked ensemble and Random Forest to predict phase stability of binary systems
- Identified 18 key features and improved product model prediction by 38%: achieved $F_1$ score 0.60

**IEEE-CIS Credit Card Fraud Detection**, Kaggle Data Challenge — Jul 2020

- Cleaned and explored transaction data of over 400 features and of high imbalance
- Trained Random Forest, LightGBM, XGBoost, and Logistic Regression models for classification
- Improved precision and recall by features selection and optimization: achieved $F_1$ score 0.71

## EDUCATION

**Ph. D. in Geochemistry**, Louisiana State University, Baton Rouge, LA — Sep 2020

**Honor**, **Leadership LSU Class**, Louisiana State University, Baton Rouge, LA — Apr 2015

**M. Sc. in Geochemistry**, Stony Brook University, Stony Brook, NY — May 2014