

Sidney Le

Data Scientist

✉ sidney.h.le@gmail.com

☎ (408) 981-9036

🌐 sidney.work

CAREER OBJECTIVE

Data scientist with experience problem-solving in many research domains, from sociology to housing to health. Strong background in statistics and skilled at communicating data and concepts. Hoping to make the world a better place.

TECHNICAL SKILLS + SOFTWARE

- **ML/DL:** scikit-learn, keras, tensorflow, pytorch, SparkML
- **NLP:** NLTK, gensim, BERT
- **Python:** numpy, pandas, matplotlib/seaborn
- **R:** tidyverse, ggplot2, shiny, rmarkdown
- **Data:** Spark, SQL
- **BI:** Tableau, Excel, PowerPoint
- **Version Control:** git, github

EDUCATION

University of California, Berkeley – Berkeley, CA
Bachelor's, Statistics
January 2015 – December 2018

Relevant coursework: Population Analysis, Advanced Linear Algebra, Probability, **Statistics**, Matrices and Graphs, **Social Network Analysis**, Computer Program Structure, Linear Modeling, Machine Learning, Time Series Analysis, Text Analysis, Collaborative Data Science

PROFESSIONAL EXPERIENCE

Data Scientist – Dascena, Oakland, CA / June 2018 – Feb 2020 (1 year as DS + 7 months as DS intern)

- Designed and implemented experiments utilizing machine- and deep-learning in **Python** to leverage large-scale clinical EHR data, including unstructured text, and drive novel health research
- Wrote and published technical papers (see below) to demonstrate novelty and significance of experimental results; developed technical aspects of grants to fund large scientific and engineering projects
- Managed and processed large-scale clinical EHR data for use in analysis using a **Linux** machine on the **AWS** cloud computing platform, **MongoDB**, and **PostgreSQL**
- Worked across teams, including engineering and sales, in order to communicate data needs and uses
- ML/DL techniques applied include: transfer and semi-supervised learning, RNNs and CNNs (implemented in Keras and Tensorflow), NLP, XGBoost

Research Associate – Goodly Labs, Berkeley, CA / Feb 2018 – Jan 2019 (1 year)

- Worked with teams of sociologists and students to develop research and social good products
- Led project development and determined technical goals and timeline
- Developed machine-learning pipeline utilizing clustering and NLP to extract sociological insight from user-generated data in R
- Managed and iterated user platform for generating data

Undergraduate Student Instructor (Data 8) – UC Berkeley, Berkeley, CA / Aug 2018– Dec 2018 (5 months)

- Taught introductory **data science** as an instructor in one the largest data science university courses in the country
- Responsibilities included: developing course material; conducting discussion and lab sections, office hours, and exam prep sessions; and advising students on data science as an interdisciplinary field of study

Data Consultant – SUSA at UC Berkeley, Berkeley, CA / Aug 2017 – May 2018 (10 months)

- Built predictive models for localized food need in collaboration with the Alameda County Community Food Bank
- Analyzed affordable housing needs/policy, developed indicators and maps for CTSP Data for Good competition, placed 2nd

SELECTED PUBLICATIONS

- Le, Sidney, et al. "**24: EFFECTS OF MONOCYTE DISTRIBUTION WIDTH AND WHITE BLOOD CELL COUNT ON A SEPSIS PREDICTION ALGORITHM.**" *Critical Care Medicine* 48.1 (2020): 12. doi: 10.1097/01.ccm.0000618596.05438.08. *Recipient of the SCCM Star Research Achievement Award.*
- Le, Sidney, et al. "**Pediatric Severe Sepsis Prediction Using Machine Learning.**" *Frontiers in pediatrics* 7 (2019): 413. doi: 10.3389/fped.2019.00413.
- Barton, Christopher, et al. "**Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs.**" *Computers in biology and medicine* 109 (2019): 79-84. doi: 10.1016/j.compbiomed.2019.04.027.