



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.

PROFESSIONAL EXPERIENCE

Data Scientist

Dascena, Oakland, CA / Jan 2019 – Feb 2020

- Plan and implement experiments utilizing machine- and deep-learning in **Python**, utilizing large-scale clinical EHR data including unstructured text, to drive novel health research
- Write and publish technical papers to demonstrate novelty and significance of experimental results; develop technical aspects of grants to fund large scientific and engineering projects
- Manage and process large-scale clinical EHR data for use in analysis using the **AWS** cloud computing platform, **MongoDB**, and **PostgreSQL**
- Work 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**), **NLP**, **XGBoost**

Research Associate

GoodlyLabs, Berkeley, CA / Feb 2018 – Jan 2019

- Worked with teams of sociologists and students to develop research and social good products
- Lead 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

TECHNICAL SKILLS + SOFTWARE

- **Python (fluent)**: **scikit-learn**, **keras**, **tensorflow**, **numpy**, **pandas**, **matplotlib**
- **R (fluent)**: **keras**, **ggplot2**, **plotly**, **shiny**
- **NLP**: **NLTK**, **gensim**

EDUCATION

University of California-Berkeley
Berkeley, CA

B.A. Statistics (Dec 2018)

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.combiomed.2019.04.027.