JULIE LAI

(224)343-2446 ⋄ julielai@berkeley.edu ⋄ https://github.com/juliejlai/data-science

EDUCATION

University of California, Berkeley
M.S. in Information and Data Science — August 2020 - August 2021
College of Letters and Science, B.A. in Physics and Astrophysics — August 2016 - May 2020

WORK EXPERIENCE

Data Scientist Intern — May 2021 - Present
Springbuk, Inc. — *Indianapolis, Indiana*
- Creating a machine learning model in Python to predict risk of Opioid Use Disorder using insurance health claim data
- Querying PostgreSQL databases, joining, and cleaning data for predictive modeling

Bioinformatics Intern — May 2019 - June 2020
Finkbeiner Lab, Gladstone Institute of Neurological Disease — *San Francisco, California*
- Refined semi-supervised machine learning models with Python to predict ALS using patient genomic data
- Visualized trends in patient genomic data using matplotlib and seaborn graphs
- Compacted large amounts of patient genomic data into a concise dataframe of patient versus variants

Astrophysics Research Assistant — February 2018 - February 2019
Nugent Research Group — *Berkeley, California*
- Categorized behavior supernovae light curves to understand causes a supernova's final explosion
- Fitted H-alpha peaks to Gaussian curves using Python to determine significance compared to noise

PROJECTS

Formality Neural Machine Translation — April 2021
- Implemented a word-level LSTM sequence-to-seqeunce Neural Machine Translation models in Keras Tensorflow to translate formality using the Grammarly's Yahoo Answers Formality Corpus
- Evaluated results using a Text Formality Classifier and visualizing formality using LIME

Predicting Alzheimer's — November 2020
- Used a Kaggle Dataset of MRI Segmentation Images to classify different stages of Alzheimer's dementia
- Classified images using a variety of models including Keras, Naive Bayes, and SVC
- Filtered the data through dimension reduction

Modeling Stellar Spectra — May 2020
- Built a generative model to predict what a stellar spectrum should look like for a given set of stellar properties using linear models and MCMC optimization
- Inferred the properties of stars by fitting their spectra and visualizing properties using matplotlib and seaborn

LEADERSHIP

Media Director — January 2019 - May 2020
The STEMinist Chronicles — *Berkeley, California*
- Promoting representation of women in STEM by executing the photoessay pipeline: interviewing, transcribing, editing, and publishing photoessays on social media platforms

STRENGTHS

Technical	Python (numpy, pandas, scikit-learn, matplotlib, seaborn, Altair), R, SQL, JavaScript (D3, Vega-lite), Tableau, Excel, LaTeX, Bash, LabView
Graphic Editing	Adobe Photoshop, Adobe Lightroom, Adobe Premiere Pro, Adobe Illustrator, Procreate
Languages	Conversational Chinese and Spanish