# FISHER MORITZBURKE

SOFTWARE ENGINEER

fmoritzburke@gmail.com
snisher.github.io (blog)
(415) 342-1513
San Francisco, CA
fisher-moritzburke-536b01162
snisher


## SKILLS

**GENERAL:**
Programming (Julia, Python), Linux
**SOFTWARE LIBRARIES:**
DataFrames.jl, CSV.jl, Query.jl, Matplotlib, Plots.jl, Test.jl
**DATABASES:** PostgreSQL, MongoDB

## EDUCATION

### University of California at Santa Cruz

BS Neuroscience 2019
Minor Computer Science
Relevant Coursework: database systems, data mining & web scraping, data structures, applied statistics, machine learning & data mining, probability theory

## PROJECTS

### ALCOHOL SALES CLUSTERING

- Analyzed 3 GB of line-item liquor sales data from Iowan liquor stores.
- Clustered products into similar categories using custom implemented k-modes algorithm.
- Tested functionality with test file.

## EXPERIENCE

**Metis** Mar. 2020 to June 2020 · San Francisco, CA
Data Scientist

Completed a 12 week full-time data science bootcamp focused on developing skills in Python, Machine Learning, Statistical Modelling and Data Visualization. Designed, implemented and presented five end-to-end data science projects demonstrating concepts such as data acquisition, supervised and unsupervised learning, classification, NLP and deep learning.

**Predicting Which Genes are Associated with Diseases**

- Trained a graph neural network using PyTorch Geometric to predict whether a gene is associated with a disease, formulated as link prediction problem on a heterogeneous graph (built with NetworkX) of 519 diseases and 7,294 genes.
- Processed clinical descriptions of diseases and genes into feature vectors using sklearn's TF-IDF vectorizer after removing stopwords.
- Disease-gene associations help focus research into new drugs, reducing costly experimental validations.

**ADHD Diagnosis Prediction from Functional Brain Connectivity Matrices**

- Trained a soft-voting ensemble classifier (logistic regression, SVM with RBF kernel, and naive Bayes) to predict whether someone will be diagnosed with ADHD based on brain connectivity.
- Calculated network statistics (degree, clustering coefficient, and closeness) with NetworkX to use as features.
- Determined feature importance using SHAP, which provided physiological insights about which brain connections are associated with ADHD.

**Presidential Speech Analysis**

- Processed presidential speeches with spaCy and analyzed topics using Gensim's LDA and CorEx.
- Stored data in MongoDB on an AWS EC2 instance.