

## KATHERINE B. SHAKMAN, PhD

773 882 8262 | San Francisco, CA | [katieshak@gmail.com](mailto:katieshak@gmail.com)

[github.com/katieshakman](https://github.com/katieshakman) | [linkedin.com/in/katherine-shakman](https://linkedin.com/in/katherine-shakman) | [katieshakman.github.io](https://katieshakman.github.io)

---

### SKILLS

**Programming Languages and Tools:** Python\*, R, MATLAB\*, MongoDB\*, SQL, PyMongo, Bash, Pandas\*, Numpy, Ray, Spark, Seaborn, Scikit-learn\*, Torch, Tensorflow/Keras, AWS\*, Unix\*, Docker, Kubernetes, Git

**Scientific:** Experimental design\*, statistical analysis\*, data visualization\*, technical communication\*

**Machine Learning and Algorithms:** Linear and logistic regression\*, decision trees and random forest\*, k-nearest neighbors, support vector machines, k-means clustering, principal components analysis and dimensionality reduction, neural networks, natural language processing, collaborative filtering

\* = most experience

### EXPERIENCE

#### Evidation Health

Senior Health Data Scientist

San Mateo/San Francisco, CA  
August 2021 – December 2022

- Collaborated on study design for research projects using wearable devices and real world data
- Designed, authored, and executed complex statistical analysis plans for sponsored studies, coalescing input from various stakeholders to synthesize business objectives into an effective scientific approach
- Developed and integrated with data pipelines to query, clean, featurize, and model data from multiple sources in a variety of formats using SQL, Python, and R
- Created compelling visualizations of results and communicated scientific and statistical methods, strategies, and results with both internal and external stakeholders in a variety of roles

#### Domino Data Lab

Field Data Scientist

San Francisco, CA  
May 2020 - August 2021

- Co-conceived, piloted, and built out the data science advisory team within professional services
- Collaborated on implementation and optimization of a distributed deep learning model for tumor detection in whole slide images, resulting in a conference talk at Nvidia GTC 2020
- Contributed design and implementation of a solution for online machine learning from streams

Field Engineer/Customer Success Engineer

October 2018-May 2020

- Educated data science leaders and teams in best practices on Domino data science platform
- Engineered solutions to customers' technical issues, from environment design to model deployment
- Produced tools to enhance the efficiency of the customer success team

#### Insight Data Science

Health Data Science Fellow

San Francisco, CA  
June 2018 - September 2018

- Consulted for Therachat to build TheraPulse, a tool for predicting patients' symptom severity
- Produced an end-to-end solution from querying a MySQL database to modeling to webapp
- Engineered 8 text-derived features using NLP tools in Python
- Predicted patient distress ratings via SVM and random forest models with grid search HPO in sklearn

#### Columbia University

Graduate Research Fellow in Neuroscience

New York, NY  
August 2011 - May 2018

- Led a team that assessed olfactory behavior and neural responses to olfactory stimuli to uncover new roles for dopaminergic neurons in modulating innate and learned responses to odors
- Wrote automated MATLAB tools to analyze neural activity in insect dopamine neurons and insect behavior in response to chemosensory stimuli

### EDUCATION

#### Columbia University

PhD in Neurobiology and Behavior

New York, NY  
May 2018

#### California Institute of Technology (Caltech)

BS in Biology

Pasadena, CA  
May 2011