MATTHEW XIE

mkxie@berkeley.edu (925) 997-1661 github.com/mkxie linkedin.com/in/matthewkxie/

EDUCATION

University of California, Berkeley **August 2017—May 2021**
BA Computer Science **|** ***GPA***: 3.73
Coursework: Data Structures, Algorithms, Databases, Operating Systems, Discrete Math and Probability Theory, Data Science, Machine Structures, Designing Information Systems and Devices
Clubs and Activities: Pioneers in Engineering (PiE), Berkeley Engineers and Mentors (BEAM), AFX Dance

EXPERIENCE

Data Science Intern—*Sandia National Laboratories, Livermore, CA* **May 2019—August 2019**
- Utilized Python and Apache Spark on Sandia's High Performance Data Analytics clusters to scale up distributed data processing on machine learning tasks and demonstrate big data applications for new users
- Built a convolutional neural network classifier with Spark and Keras for deep learning on tens of thousands of seismic records to distinguish earthquakes from man-made quarry blasts with 97% accuracy and gain actionable insights for future inferences
- Wrote tutorials and guides to introduce new users to machine learning workflows using Spark and Keras on Sandia's data analytics clusters, including setting up Spark, configuring virtual environments for both CPU and GPU architectures, distributing work, and monitoring job status

Engineering Intern—*Lawrence Livermore National Laboratory, Livermore, CA* **June 2018—August 2018**
- Automated installation and configuration processes on Linux-based HPC clusters using Ansible (cmd/Bash)
- Worked with a team of four to add new functionality to a Python software suite based off of the open-source Pyvmomi library for virtual machine automation and orchestration for LLNL's Livermore Computing and Systems Administration Group (Python)
- Improved efficiency by eliminating the need to do tasks through VMware's ESXi GUI, and presented to industry experts for feedback

Software Developer—*Pioneers in Engineering, Berkeley, CA* **September 2018—May 2019**
- Implemented an internal staff check-in database to keep track of attendance during worksessions using Python, SQLite, and Google Cloud's NoSQL-based Datastore
- Maintained and expanded features for student worksessions web app built with Ruby on Rails using Google Sheets API, Slack API, Flask, and Python to check staff availability and send daily Slack notifications

Data 100 Academic Intern—*UC Berkeley Division of Data Sciences, Berkeley, CA* **January 2019—May 2019**
- Taught students challenging concepts of data science such as Pandas, SQL, classification, and regression during labs, homeworks, and projects

PROJECTS

Concurrent Cached Filesystem (Go) **April 2019—May 2019**
- Implemented Go concurrency on a cached file system for multithreaded execution of file reads from disk and placing them in the cache for faster file handling

Spam Email Classification (Python) **October 2018—November 2018**
- Constructed a logistic regression model using Pandas and Scikit-learn with NLP techniques to classify spam vs ham emails with 90% accuracy

Bay Area Maps (Java) **April 2018**
- Provided the back end for a Google Maps-like application for the East Bay Area, including zoom capabilities, rendering map images from user queries, and finding shortest routes using A* search algorithm

SKILLS

Languages: Python, Java, C, Golang, HTML/CSS, Ruby/Rails, SQL
Technologies: Git, NumPy, Pandas, Spark, TensorFlow, Keras, Flask, SQLite, Docker, Linux, HPC