Scott Lee

Data Scientist / Software Engineer

Email scott.lee.3898@berkeley.edu
Website scottjlee.github.io
Location Berkeley, CA
Phone 408-499-4014

Education

University of California, Berkeley — B.A. Computer Science *2020** — GPA: 3.85 / 4
(*B.A. CS 2019 + M.S. CS 2020 or B.A. CS + B.A. Statistics 2020, pending acceptance to M.S. program)
Selected Coursework: Algorithms, Data Structures, Machine Learning, Artificial Intelligence, Database Systems, Computer Architecture, Discrete Math, Probability & Random Processes, Statistical Theory, Linear Modeling

Experience

Software Engineering Intern   Rubrik — Palo Alto, CA   May 2018 - Aug 2018

— Designed and implemented a robust database system with efficient random reads and sequential scanning, used to store email metadata in the cloud.
— Added universal compatibility with various cloud providers (AWS, Azure, Google Cloud) and further augmented it with Zstd compression and file deduplication.

Research Assistant   RISELab — Berkeley, CA   Feb 2018 - Present

— Applied convolutional neural networks to detect and classify arrhythmias from electrocardiogram readings.
— Also built a hierarchical classifier that parses and analyzes data from doctors' notes.

Head Teaching Assistant (UGSI)   Data 100 — Berkeley, CA   Aug 2017 - Present

— Prepared materials, taught lab section, held office hours, and developed assignments.
— Facilitated class logistics, including managing staff, organizing exams, and overseeing enrollment for a class of 1400 students.

Data Science & Education Intern   Brilliant — San Francisco, CA   May 2017 - Aug 2017

— Conducted analysis using SQL and Python on 1+ million emails to evaluate email algorithm effectiveness and created a proposal that increased clickthrough rate by 30%.
— Designed and produced curricula for "Math for Computer Science" and "Intro to Data Science" courses.

Technical Skills

Python (+ Pandas, Scikitlearn, TensorFlow, Apache Spark), SQL, R, Java, Go, C
HTML, CSS/LESS, Javascript, jQuery, Bootstrap

Projects (More on my website; links to live project are boxed.)

BerkeleyTime

— BerkeleyTime is an augmented course catalog that provides data on courses, enrollment trends, grade distributions, and more. I serve as the Product Manager and Data Scientist for BerkeleyTime, and was previously Lead Engineer.
— We just wrapped up a course scheduler, and are currently working on a data-centered initiative with student course data (such as course recommendations, automatic course plan generation, and intelligence course classification).

Neural Net Applications — Python

— Implemented a neural net from scratch (graph class, implementing matrix operations, backpropagation, etc).
— Applied neural nets to solve various problems, such as digit classification, language classification, and function approximation.

Song Classifier — Python, Tensorflow

— Created a deep learning classifier that categorizes songs as either country or hip-hop based on lyrics frequencies.
— Achieved an accuracy rate of 89% and was selected as one of 10 winners of a 200-member Kaggle contest.

Jesture (Cal Hacks 3.0) — HTML, CSS, JS, jQuery, Bootstrap, JSON, Python

— Built a utility application that detects different gestures and triggers varying options (e.g. pause music on Spotify).
— Implemented API for Spotify, Slack, Facebook, and more, then linked to a sleek web UI.