KEVIN KAO

San Jose, CA 95132 | (408) 373-5520 | kkao@berkeley.edu
github.com/kk415kk

EDUCATION

University of California, Berkeley May 2016
B.S. Electrical Engineering & Computer Science 3.63 – CS GPA
Awards: Regents & Chancellor Scholar, Dean's Honor, ACM Publication

My areas of interest include **algorithms**, **applied machine learning**, **systems design**, **search relevance**, and **NLP**.

PUBLICATIONS

[1] Zachary Pardos, Kevin Kao. "moocRP: An Open-Source Analytics Platform". March 2015. Learning @ Scale 2015.
[2] Kevin Kao, Ananth Sub., Valerie Liu. "Machine Learning with fMRI Brain Images." December 2014. CS194-16 BIDS Poster Session.

SKILLS

Proficient with Java, Python, JavaScript, HTML/CSS, UNIX, Git, jQuery, SQL, MySQL, Redis, nginx
Experience with Scala, Node.js, Sails.js, Flask, Ruby on Rails, C++, Bash, SVN
Data Analytics Experience: Hadoop (Pig, MapReduce, HBase), R, Spark, numpy, scikit-learn, scipy, OpenCV, OpenNLP, Pandas, Encog

WORK EXPERIENCE

Software Engineer, SEO/Search September 2015 – Present
LinkedIn
- Designing and implementing new processing pipeline in LinkedIn's infrastructure for search snippets generation and storage in production, which involves various components such as Kafka, Hadoop, Couchbase caches, etc

Software Engineering Intern, SEO/Search May 2015 – August 2015
LinkedIn
- Developed **search snippets library** for job search results using **NLP and deep learning machine learning** techniques
- Designed entire workflow from scratch, from initial research to data collection design to algorithmic implementation into a new library

Machine Learning Research Intern February 2014 – May 2015
UC Berkeley School of Information
- Developed **Node.js web app** called moocRP that serves as a data distribution and data analytics sharing platform for researchers
- Contributed significant data transformation Python and Bash scripts to **open-source Stanford/Harvard data analytics scripts**
- Designed **analytics sharing system** that allows users to upload D3 visualizations to be automatically scaffolded and shared publicly
- Implemented **secure datasets download pipeline** with SSL, CSRF-protection, CAS authentication module, GPG encryption
- Applying deep learning techniques such as convolutional neural nets, LSTMs, and spectral clustering on MOOC data

Software Engineering Intern, Systems August 2014 – January 2015
LogicBlox
- Researched cutting-edge academic papers on implementation of **dynamic SCC algorithm** under edge additions/deletions
- Designed and implemented internal graph structure optimized for **incremental graph maintenance**, avoiding full recomputes of SCCs

Software Engineering Intern, Data/Infrastructure May 2014 – August 2014
AutoGrid Systems
- Fully developed new vital portion of data analytics pipeline: an **incremental ETL algorithm** using multithreading, the star schema, Redis, MySQL, and MapReduce to allow scalable incremental computations on millions of data points by ML algorithms
- Developed a **communications platform** (email, SMS, voice) Ruby gem with an enhanced templating system as an infrastructure service
- Developed **internal monitoring app** with Python Flask to check config/status of applications, ensuring quick detection of problems

PROJECTS

Image Algorithms (Python, scipy, numpy) September 2015 – Present
- Implemented image alignment algorithms for colorizing images, and optimized for high res images through use of image pyramids

fMRI Brain Image Analytics (Python, Scala, scikit-learn, skimage, OpenCV, numpy) October 2014 – December 2014
- Performed exploratory data analysis on fMRI images to classify voxel responses based on features drawn on stimulus images
- Developed classification, clustering, and regression models to predict brain responses, as well as data wrangling processing algorithms
- Yielded significant results with binary classification of brain region activity, with approximately 90% accuracy and on par with baseline

SimpleDB Relational Database System (Java) January 2014 – May 2014
- Designed a transactional database that allows concurrency at a page level, with locking to prevent race conditions
- Wrote sort-merge algorithm and DP query optimization algorithm that generated close-to-optimal query plans