## Pei Yong Sim

UC Berkeley EECS student looking for full-time opportunities in software engineering (w/ focus in machine learning) after **graduating in May 2017**.
All project code, documentation & **demo** available at www.pysim.me/projects.

Berkeley, CA, 94704
**(831) 428-3525**
**py@pysim.me**
www.pysim.me


## EXPERIENCES

### Synocate, Palo Alto — *Software Engineer Intern*

July 2016 - Sep 2016

Role: Helped build tools to enhance college admissions experience.

**Projects**
1. **Web scraper and browser emulator** to gather essay prompts to be populated to our site (tech stack include Node.js, Express, Swagger, PostgreSQL).
2. OkCupid's **matching algorithm** to best match counselors with students which greatly improved customer satisfaction.
3. Summer programs **recommender system** prototype w/ basic keyword search & results ranking based on user profiles.

### University of California, Berkeley — *Academic Intern*

Jan 2016 - Present

Role: Help students in CS 189 (Machine Learning) and CS 61A (Intro to CS) with homework, projects and labs. Also help hold review sessions.

## SELECTED PROJECTS

**Systems/Databases**
**1. Compiler** — Compiles an Object-Oriented language source code into Java bytecode like language and then executes the compiled class files.
**2. DBMS** — SQL (relational) database management system that supports CRUD & Join operations, B+ trees indexing to enhance performance, query optimization and concurrency control. Written in Java.
**3. NoSQL** — Distributed Key Value Store that uses 2PC protocol for leader & follower servers coordination. It is fault tolerant and supports crash recovery.
**4. PintOS** — Unix operating system framework that supports multi-threading, user programs execution and file systems w/ buffer cache.
**5. PyChat**—A full-stack (MVC) web application that supports user management, instant messaging (WebSocket on TCP/IP) & A.I. ChatBot. Also provides REST API.

**A.I. & Machine Learning**
**1. Netflix** — Movie recommender system using latent factor analysis.
**2. OCR** — Neural network that does handwritten digit recognition w/97% accuracy
**3. PySpam** — Spam Classifier implemented using Decision Trees.
**4. Supply Chain Management** — Use unsupervised learning to help a wholesale distributor identify the best delivery service to meet the needs of each customer.
**5. Sentiment Analysis** — Simple Natural Language Processing application.

## EDUCATION

### University of California, Berkeley - *B.S. in EECS*

Graduating in May 2017, GPA: 3.2

**Coursework:**
**1. Fundamentals:** Data Structures, Algorithms, Discrete Math, Systems Design I/II
**2. Systems:** Computer Architecture, Operating Systems, Computer Networking, Compilers, Databases
**3. Data Analytics**: Data Science, Artificial Intelligence, Machine Learning, Probability in EECS, Game Theory

## SKILLS

**Programming Languages**
1. Proficient: Python, Java, C/C++, SQL
2. Familiar: JavaScript, Node.js

**Frameworks**
1. Distributed Computing: Spark (w/ MapReduce)
2. Web: Flask, Express, HTML, Bootstrap CSS

**Libraries:**
1. Data Analytics: TensorFlow, Sklearn, Numpy, Matplotlib
2. Testing: JUnit
3. Web: jQuery, Scrapy

**Other tools:**
Git, Maven, Heroku Cloud, AWS, PostgreSQL, mongoDB

## EXTRACURRICULARS

**Data Science Society at Berkeley**
Social Network Analysis on borrowers and lenders on Kiva's microlending platform.

**Math Tutor** Provided tutoring assistance to students at the Math Learning Center.

**Hack UCSC** Semi-finalists in the hackathon where my team and I built cccPlan using Node.js and mongoDB.

## LANGUAGES

Chinese, Cantonese, Malay