Edward Gan

egan1@stanford.edu edgan8.github.io

Interested in connecting data and product, with experience in both data modeling and engineering.

Education

- **Stanford University** Stanford, CA
  PhD Student, Computer Science *2015 - Present*
  - Research with Prof. Peter Bailis on designing practical machine learning systems
  - Developed unsupervised classification and feature selection algorithms for monitoring applications
  - Deployed anomaly explanation algorithms at local tech companies and diagnosed production issues
- **Stanford University** Stanford, CA
  M.S. in Computer Science, GPA 4.1/4.0 *Sep 2017*
- **Harvard University** Cambridge, MA
  A.B. Summa Cum Laude in Computer Science and Mathematics *May 2013*

Experience

- **Airbnb** San Francisco, CA
  Price Modeling Engineering Intern *Summer 2016*
  - Set up machine learning estimates in Spark to target users for customized price suggestions
  - Targeted suggestions achieved 20% conversion rate over tens of thousands of hosts
  - Conducted end-to-end variability analyses on the pricing pipeline to prioritize model improvements
- **Facebook, Inc.** Menlo Park, CA
  Software Engineer, Data Tools + Ads Data Targeting *2013 - 2015*
  - Data Tools
    * Re-built the execution engine and UI for backfilling Hadoop pipelines on historical data
    * Iterated with business intelligence users to design interface that supports ad-hoc workflows
  - Offsite Ads Targeting Backend
    * Defined metrics, real-time dashboards, and remediation to reduce outages from hours to minutes
    * Built k-means ML pipeline used by marketing to segment ads audiences

Selected Publications

- **Scalable Kernel Density Classification via Threshold-Based Pruning** ACM SIGMOD
  Edward Gan, Peter Bailis *May 2017*
  - An unsupervised anomaly detection algorithm that outperforms scikit-learn by orders of magnitude on high-dimensional data.
- **MacroBase: Prioritizing Attention in Fast Data** ACM SIGMOD
  P. Bailis, E. Gan, S. Madden, D. Narayanan, K. Rong, S. Suri *May 2017*
- **Expanding mutually exclusive clusters of users of ...** Patent Pending US20170024455A1
  S. Powell, B. Arnoux, S. Hong, D. Chapsky, E. Gan, et al *July 2015*

Skills + Awards

- Proficient with Python, Java, Pandas, and SQL/Hive. Familiar with Scikit-learn, Spark, R, and C++.
- NSF Graduate Research Fellowship 2015-2020. International Physics Olympiad Gold Medalist 2008.