

same
Eric Doi

<https://www.linkedin.com/in/ericdoi/>
<https://www.linkedin.com/in/ericdoi/>

Email : <mailto:eric.kosei.doi@gmail.com>

Mobile : +65 8768-5351

EXPERIENCE

- **Grab** - Southeast Asia's leading ride-hailing/deliveries/payments platform Singapore
Lead Data Scientist *Aug 2018 - Present*
 - **Map search relevance:** Spearheaded the optimization of our map search system in collaboration with engineering, product, and operations teams; set improvement roadmap, established performance metrics, standardized evaluation pipelines, tuned parameters in Elasticsearch and LTR models, and executed analyses and A/B tests.
 - **Growth-hacking:** Optimized video content using popularity heuristics for ~4x lift in views; prototyped NLP-based interest graph in Neo4j; used Locality-Sensitive Hashing to identify ~5x increase in similarity graph edges.
- **Retention Science** - A SaaS startup for intelligent marketing Santa Monica, CA
Data Scientist *Apr 2015 - May 2018*
 - **Lead scoring:** Overhauled SparkML model for customers' likelihood of purchase; used regularization, feature engineering, probability calibration, and operational logic to improve PR-AUC by 20x and secure client renewal.
 - **Product recommendations:** Wrote Scala-Spark pipelines to transform clicks and transactions into item recommendations to 100M+ users daily; e.g. used Bloom Filters for 10x speedup of duplicate removal.
 - **A/B testing service:** Created a Play/Spark service to accelerate testing of ML models in production.
 - **Subscription churn prediction:** Built a Random Forest classifier in H2O with AUC of 0.8 to aid churn prevention.
 - **Interim management:** Led team of 5 for 1y, triaging feature requests and upholding code standards.
- **Opera Solutions** - Enterprise analytics software and consulting firm San Diego, CA
Analytics Manager (2014) / Analytics Lead (2013) / Sr. Analytics Specialist (2011) *Apr 2011 - Apr 2015*
 - **Healthcare billing optimization:** Ran modeling, maintenance, and analysis for a live system (Patent 9,785,983) used by ~30 auditors to recover ~\$10M/year revenue by recommending missing HCPCS charge codes.
 - **Feedback model:** Introduced a second-layer model to learn user feedback and cut false positives by ~30%.
 - **Model deployment:** Trained, deployed, and evaluated models for 100+ hospitals, with mean of ~40% precision.
 - **Management:** Led Delhi-based analytics team of 6, collaborated with software team, and presented to clients.
- **Laserfiche** - Pioneer in document imaging system software since 1987 Long Beach, CA
Software Engineer Intern *Jun 2010 - Sep 2010*
 - **Mobile document scanner:** Ported a de-warping algorithm (Canny edge detection and Hough transform) from desktop to iPhone 3G; used heuristics to maintain speed and accuracy with 94% less RAM and 80% less CPU.

EDUCATION

- **University of California, San Diego** La Jolla, CA
MS in Computer Science (Artificial Intelligence); GPA: 3.9 *Sep 2009 - Mar 2011*
 - **Capstone:** *Low-Rank Decomposition and Logistic Regression Methods for Link Prediction in Terrorist Networks*
- **Harvey Mudd College** Claremont, CA
BS in Computer Science; GPA: 3.5, with Distinction *Aug 2005 - May 2009*
 - **Research:** *Grouping Strokes into Shapes in Hand-Drawn Diagrams*, AAAI 2010
 - **Awards:** Harvey S. Mudd Merit Scholar; Harvey Mudd National Merit Scholar

OTHER

- **Kaggle:** Top 10% rank in rental listing classification competition (github.com/ericdoi/kaggle-renthop-2017https://github.com/ericdoi/kaggle-renthop-2017).
- **Eagle Scout:** As Senior Patrol Leader, led a troop of 40+ scouts in camping, community service, and growth.
- **Interests:** Sustainable energy, Mandarin, cognitive psychology, armchair economics, rock climbing.

SKILLS

- **Software:** Languages (Scala, Python, C*, Bash), Tools (git, JIRA), Infra (AWS, Azure)
- **Modeling:** Packages (SparkML, scikit-learn, numpy/scipy, R, H2O), Algorithms (Linear/Logistic Regression, Trees, Stacking; Collaborative Filtering; PCA/SVD, K-Means; Uplift modeling; Reinforcement learning)
- **Data:** ETL (Spark, Airflow, Parquet, Oozie), Analysis (SQL, pandas, Excel, cut/grep/awk, dplyr)