Jingyi Kenneth Tay

kjytay.github.io
kjytay@stanford.edu | 650-250-3790

linkedin.com/in/kjytay/

EDUCATION

STANFORD UNIVERSITY | Ph.D. in Statistics
Expected Jun 2021 | Stanford, CA
Advisor: Robert Tibshirani.

PRINCETON UNIVERSITY | A.B. in Mathematics
Jun 2010 | Princeton, NJ
Summa Cum Laude • Certificates in Program of Applied & Computational Mathematics, Program of Finance
Senior Thesis Advisor: Ramon van Handel.
Junior Independent Work Advisor: Robert Calderbank.

WORK EXPERIENCE

GOOGLE | Data Scientist Intern (Payments Data Science)
Jun 2020 – Sep 2020 | Sunnyvale, CA
- Developed novel method and R package for computing variance for post-stratified estimator in potential outcomes setting. In one application, confidence interval width was reduced by 11%.
- Developed new algorithm that reports an experiment's heterogeneous treatment effect concisely while ensuring statistical validity. This work enables analysts to quickly understand how the treatment varies along dimensions of interest.

A9.COM, AMAZON SEARCH | Applied Scientist Intern (Search Relevance)
Jun 2019 – Sep 2019 | Palo Alto, CA
- Conceptualized and constructed data pipelines for new, granular metrics for Amazon search relevance models. Processed ∼1B queries and ∼20B item responses to obtain dataset for predictive modeling.
- Built a model based on these pre-experiment metrics to predict performance on live customer traffic, so that experimental bandwidth can be allocated more efficiently. Model improved test performance metric by 20% over baseline.

STANFORD UNIVERSITY | Instructor & Teaching Assistant
Sep 2016 – Present | Stanford, CA
- Coached first-year statistics PhD students for qualification examination in applied statistics. 100% pass rate.
- Developed new material for and taught "Introduction to R" course for undergraduates 3 times. 94% of students described instruction of the course as "Excellent" or "Good".

INFOCOMM DEVELOPMENT AUTHORITY | Data Scientist (Data Science Division)
Oct 2015 – Aug 2016 | Singapore
- Spearheaded engagements with a wide array of government agencies (economic, transport, social) to analyze their data to support public policymaking. Responsibilities included project scoping, data cleaning, visualization, statistical analysis and presentation of results.
- Systematized and tested recruitment framework and materials for all roles in the division, including data scientist, quantitative strategist and front-end developer.
- Developed division's operating policy for data management and statistical disclosure control.

MINISTRY OF THE ENVIRONMENT & WATER RESOURCES | Assistant Director (Environmental Policy Division)
Sep 2013 – Sep 2015 | Singapore
- Drove progressive policies to ensure sustainability and efficiency of Singapore's waste management system.
- Evaluated usefulness of movement data in predicting spread of Chikungunya virus in Singapore.
- Chairman of Staff Well-Being Committee (Apr 2014 - Mar 2015): Led team of 10 officers in conceptualizing and executing activities to improve staff welfare and morale. Also managed and accounted for budget (20K+) for staff welfare.

MINISTRY OF DEFENSE | InfoComm Technologies Engineer
Apr 2012 - Aug 2013 | Singapore
- Evaluated operational performance of critical intelligence systems.

- Made performance more transparent by developing and implementing a new reporting dashboard for senior management.
- Strengthened in-house user adoption of systems through crafting and delivering technical presentations.

RESEARCH EXPERIENCE

PUBLICATIONS

- J. K. Tay, and R. Tibshirani. (2020). Reluctant generalized additive modeling. International Statistical Review, to appear, 2020. URL https://arxiv.org/abs/1912.01808. R package `relgam`.
- J. K. Tay, J. Friedman, and R. Tibshirani. (2020). Principal component-guided sparse regression. Canadian Journal of Statistics, to appear, 2020. URL https://arxiv.org/abs/1810.04651. R package `pcLasso`.
- D. Shung, C. Tsay, L. Laine, D. Chang, F. Li, P. Thomas, C. Partridge, M. Simonov, A. Hsiao, J. K. Tay, and A. Taylor. (2020). Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. Journal of Gastroenterology and Hepatology, to appear, 2020.
- D. Shung, E. Castro, J. Huang, J. K. Tay, M. Simonov, L. Laine, and S. Krishnaswamy. (2020). Neural network predicts need for red blood cell transfusion for patients with acute gastrointestinal bleeding admitted to the intensive care unit. Scientific Reports, to appear, 2020. URL https://www.medrxiv.org/content/10.1101/2020.05.19.20096743v1.
- D. L. Shung, B. Au, R. A. Taylor, J. K. Tay, S. B. Laursen, A. J. Stanley, H. R. Dalton, J. Ngu, M. Schultz, and L. Laine. (2020). Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. Gastroenterology, 2020, 158(1):160-7.

SOFTWARE

- Contributor to `glmnet` R package. Extended `glmnet` in v4.0 so that it can efficiently fit any generalized linear model with the elastic net penalty.

PREPRINTS AND PAPERS UNDER REVIEW

- J. K. Tay, N. Aghaeepour, T. Hastie, and R. Tibshirani. (2020). Feature-weighted elastic net: using "features of features" for better prediction. arXiv:2006.01395 [stat.ME], 2020. URL https://arxiv.org/abs/2006.01395. R package `fwelnet`.
- J. K. Tay, and R. Tibshirani. (2018). A latent factor approach for prediction from multiple assays. arXiv:1807.05675 [stat.ME], 2018. URL https://arxiv.org/abs/1807.05675.

AWARDS & HONORS

- Honorable Mention, American Statistical Association's Statistical Learning and Data Science Student Paper Competition (2019, 2020)
- Departmental Teaching Assistant Award (2017, 2018)
- Two Sigma Graduate Fellowship in Statistics (2017)
- Early Induction to Phi Beta Kappa (top 1% of cohort) (2009)
- Shapiro Prize for Academic Excellence, Princeton University (2007, 2008)
- Honorable Mention, William Lowell Putnam Competition (2006, 2008)
- Public Service Commission Overseas Merit Scholarship (Open) (full-ride college scholarship) (2006-2010)
- Sliver Medal, International Mathematical Olympiad (2004, 2005)

COMPUTER SKILLS

Proficient in:
Python, R
Familiar with:
C, C++, FORTRAN, PySpark, SQL, Tableau