Jingyi Kenneth **Tay**

kjytay.github.io | linkedin.com/in/kjytay
kjytay@gmail.com | 650-250-3790

EDUCATION

STANFORD UNIVERSITY | Ph.D. in Statistics
Jun 2021 | Stanford, CA
Jerome H. Friedman Applied Statistics Dissertation Award
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

LINKEDIN | Senior Data Scientist (Optimization, Data & Artifical Intelligence Foundations)
Nov 2022 - Present | Mountain View, CA

- Lead for the development and use of large-scale linear programming.

LINKEDIN | Senior Data Scientist (Experimentation Science, Data Science & Research Productivity)
Sep 2021 - Nov 2022 | Mountain View, CA

- Methods lead for the use of observational causal inference within Linkedin.

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 – Jun 2021 | 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**, B. Narasimhan, and T. Hastie. (2023). Elastic net regularization paths for all generalized linear models. *Journal of Statistical Software*, 2023, 106(1):1-31. R package `glmnet`.
- **J. K. Tay**, N. Aghaeepour, T. Hastie, and R. Tibshirani. (2021). Feature-weighted elastic net: using "features of features" for better prediction. *Statistica Sinica*, 2021. R package `fwelnet`.
- D. Shung, J. Huang, E. Castro, **J. K. Tay**, M. Simonov, L. Laine, R. Batra, and S. Krishnaswamy. (2021). Neural network predicts need for red blood cell transfusion for patients with acute gastrointestinal bleeding admitted to the intensive care unit. *Scientific Reports*, 2021, 11:8827.
- **J. K. Tay**, J. Friedman, and R. Tibshirani. (2021). Principal component-guided sparse regression. *Canadian Journal of Statistics*, 2021. 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. (2021). Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. *Journal of Gastroenterology and Hepatology*, 2021, 36(6):1590-7.
- **J. K. Tay**, and R. Tibshirani. (2020). Reluctant generalized additive modeling. *International Statistical Review*, 2020, 88(S1):S205-S224. R package `relgam`.
- 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.

CONFERENCES AND WORKSHOPS

- A. Gupta, S. S. Keerthi, A. Acharya, M. Cheng, B. O. Elizondo, R. Ramanath, R. Mazumder, K. Basu, **J. K. Tay**, R. Gupta. (2023). Practical Design of Performant Recommender Systems using Large-scale Linear Programming-based Global Inference. In *KDD 2023*.

SOFTWARE

- Contributor to `dualip` Scala package (Linkedin's open-source package for performing large-scale linear programming).
- Author of `cvwrapr` R package. Tools for performing cross-validation.
- Contributor to `glmnet` R package. v4.0: Extended `glmnet` to efficiently fit any generalized linear model with the elastic net penalty. v4.1: Added ability to fit stratified Cox models and Cox models for start-stop data, opening the way to fit a wide array of regularized Cox models (e.g. time-dependent covariates, left truncation, multiple events per subject).

PREPRINTS AND PAPERS UNDER REVIEW

- E. Tuzhilina, T. J. Hastie, D. J. McDonald, **J. K. Tay**, and R. Tibshirani. (2022). Smooth multi-period forecasting with application to prediction of COVID-19 cases. *arXiv:2202.09723 [stats.ME]*, 2022. URL https://arxiv.org/abs/2202.09723.
- **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

- Jerome H. Friedman Applied Statistics Dissertation Award (2021)
- 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, Scala, Spark, SQL, Tableau