Jingyi Kenneth **Tay**

✉ kjytay@gmail.com
kjytay.github.io/
kjytay
kjytay

EDUCATION

STANFORD UNIVERSITY | Ph.D. in Statistics
Jun 2021, Stanford, CA
Jerome H. Friedman Applied Statistics Dissertation Award.
Thesis: Extending the reach of the lasso and elastic net penalties: Methodology and practice.
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, Foundational Artificial Intelligence Technologies)
09/2021 - 12/2024, Mountain View, CA
- Optimization, Foundational Artificial Intelligence Technologies (11/2022 - 12/2024): Tech lead for the development and use of large-scale constrained optimization and contextual bandits.
- Experimentation Science, Data Science & Research Productivity (09/2021 - 11/2022): Methods lead for the use of observational causal inference within LinkedIn.

GOOGLE | Data Scientist Intern (Payments Data Science)
06/2020 - 09/2020, 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)
06/2019 - 09/2019, 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
09/2016 - 06/2021, 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)
10/2015 - 08/2016, Singapore
- Spearheaded engagements with a wide array of government agencies (economic, transport, social) to analyze their data to support public policy making. 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)

09/2013 - 09/2015, Singapore

- Drove progressive policies to ensure sustainability and efficiency of Singapore's waste management system.
- Chairman of Staff Well-Being Committee (04/2014 - 03/2015): 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

04/2012 - 08/2013, Singapore

- Evaluated operational performance of critical communications systems, including development and implementation of a new reporting dashboard for senior management.
- Strengthened in-house user adoption of systems through crafting and delivering technical presentations.

RESEARCH EXPERIENCE

Publications

1. C. Wei, B. Zelditch, J. Chen, A. A. S. T. Ribeiro, **J. K. Tay**, B. O. Elizondo, K. Selvaraj, A. Gupta, and L. B. De Almeida. Neural Optimization with Adaptive Heuristics for Intelligent Marketing System. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024: 5938-5949.
2. E. Tuzhilina, T. J. Hastie, D. J. McDonald, **J. K. Tay**, and R. Tibshirani. (2023). Smooth multi-period forecasting with application to prediction of COVID-19 cases. *Journal of Computational and Graphical Statistics*, 2023.
3. **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`.
4. **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`.
5. 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.
6. **J. K. Tay**, J. Friedman, and R. Tibshirani. (2021). Principal component-guided sparse regression. *Canadian Journal of Statistics*, 2021. R package `pcLasso`.
7. 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.
8. **J. K. Tay**, and R. Tibshirani. (2020). Reluctant generalized additive modeling. *International Statistical Review*, 2020, 88(S1):S205-S224. R package `relgam`.
9. 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

1. 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

1. Contributor to `dualip` Scala package (Linkedin's open-source package for performing large-scale linear programming).
2. Author of `cvwrapr` R package. Tools for performing cross-validation.
3. 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

1. **J. K. Tay**, and R. Tibshirani. (2018). A latent factor approach for prediction from multiple assays. *arXiv:1807.05675 [stat.ME]*, 2018.

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 Honor Society (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)
- Silver Medal, International Mathematical Olympiad (2004, 2005)

COMPUTER SKILLS

- Proficient in: Python, R, Scala, Spark
- Familiar with: C, C++, SQL, Tableau