Yinpu Li

Address: Seattle, WA, 98101 | Phone: 202-344-0121 | Email: yinpuli@icloud.com
LinkedIn: https://www.linkedin.com/in/yinpuli | Website: yinpuli.github.io |
Google Scholar: Yinpu Li | ORCID: https://orcid.org/0000-0002-4807-5541

EDUCATION

PhD in Statistics, Florida State University, Tallahassee, FL 08/2017 – 12/2021
MS in Statistics, George Washington University, Washington, DC 08/2015 – 05/2017
BS in Statistics, East China Normal University, Shanghai, China 09/2011 – 06/2015

SKILLS

MLOps / Cloud Computing: Proficient in deploying, managing, and version controlling machine learning models in production using Python, R, and Julia. Hands-on experience with AWS, Docker, Git, and parallel computing.
Algorithms: Skilled in Bayesian models (Pyro), high-dimensional and generalized classification and regression (scikit-learn, statsmodels), ensemble-trees (CatBoost, XGBoost, LightGBM, Random Forest), deep learning (PyTorch, TensorFlow-Keras), reinforcement learning, and multi-armed bandits (RL-Coach, PyQ-Learning, Vowpal Wabbit, SMPyBandits), experiment design, optimization, A/B testing, causal inference (causalinference, Causallib), survival models, longitudinal data, and stochastic modeling (PyMC3, SimPy).
Communication: Ability to effectively communicate technical ideas to both technical and non-technical stakeholders in a collaborative manner.

EXPERIENCE

Senior Data Scientist / MLE 03/2023 – Present
Supply Chain Division, Forecasting & User Tool, Chewy, Bellevue, WA

- **Anomaly Detection (LSTM/EM-GMM)** Leading a dynamic anomaly detection initiative targeting irregularities in sales, customer behavior, and inventory levels across the network; investigating root causes and collaborating with cross-functional teams, and developed concise visualizations and reports for actionable insights.

Supply Chain Division, Inventory Management, Chewy, Bellevue, WA

- **Substitutable SKUs (NLP)** Developed scalable Natural Language Processing (NLP)-driven solutions to optimize product searching / matching and recommending substitutable SKUs, reducing unnecessary SKU acquisitions, and enhancing cost-effectiveness.
- **Outbound Cost (Hierarchical Regression)** Devised an scalable hierarchical model for forecasting outbound shipping costs, extending its predictive capability to untrained routes, resulting in a significant enhancement in accuracy from MAE = $10.52/order to MAE = $1.58/order (with weighted MAPE = 5.66%).
- **Inventory Optimization (Multi-Objective DP)** Improved the current optimization system by accurately capturing the cost and benefits of transferring inventory, considering various factors such as item characteristics, network topology, customer behavior, and supplier availability.

Research Scientist 11/2021 – 03/2023
Carrier Experience & Marketplace Research Science & Analytics Org, Convoy Inc, Seattle, WA

- **Precise Matching (Multi-state Markov)** Designed and deployed a multi-state Markov model for personalized trucker cost optimization, differentiated customers' expectation and guaranteed services, resulting in a 116bps increase in on-time service and a 2.4% reduction in failure to match. The project delivered an annual cost savings of $7M.
- Full-stack responsibilities included project framing and scoping, data preparation and engineering, ML pipeline deployment and maintenance, and design of monitoring and tracking tools (Datadog, Metabase). Conducted A/B testing, statistical inference, and pattern mining to drive business insights.

Statistical and Quantitative Science Data Sciences Intern 06/2021 – 08/2021

Takeda Pharmaceutical, Boston, MA

- **Clinical Trials / Phase III Prediction (Bayesian Ensemble Tree)** Developed a novel approach that blended statistical methods and machine learning techniques to enhance predicted patient-level survival outcomes in Phase III clinical trials using external studies. Resulted in a book chapter.

Mathematical Sciences Graduate Intern 06/2020 – 08/2020

Statistical Engineering Dept, National Institute of Standards and Technology, ORAU/ORISE, Gaithersburg, MD

- **Uncertainty Pyramids in Forensics (Stats Uncertainty & Optimization)** Implemented a model uncertainty evaluation tool (R packages) for statistical algorithms in violent crime and forensic investigations, following the framework established by NIST scientists. The developed tool was recognized and highlighted on the website https://orise.orau.gov/nsf-msgi/profiles/li.html.

Research Assistant 01/2020 – 05/2020

Department of Statistics, Florida State University, Tallahassee, FL

- Developed Bayesian Nonparametric methodologies and theories by proposing new tuning-parameter-free machine learning methods for density regression and survival analysis, capable of automatically conducting feature selection, and are robust to model assumption violations. Resulted 2 top journal papers in JASA and Bayesian Analysis.

Teaching Assistant 09/2018 – 05/2019

Department of Statistics, Florida State University, Tallahassee, FL

- Independently taught intro-level applied statistics courses, Business Statistics and Statistics via Examples; course topics cover: sampling methods, statistical visualization, distribution theory, hypothesis test and linear regression.

SELECTED PUBLICATIONS

- **Li, Yinpu**, Antonio R. Linero, and Jared S. Murray. "Adaptive Conditional Distribution Estimation with Bayesian Decision Tree Ensembles." *Journal of the American Statistical Association (2022): 1-14*. *(Awarded for NSF: Bayesian Tree Models for Next-Generation Studies in the Behavioral and Social Sciences)*
- Linero, Antonio R., Piyali Basak, **Yinpu Li**, and Debajyoti Sinha. "Bayesian Survival Tree Ensembles with Submodel Shrinkage." *Bayesian Analysis 1, no. 1* (2021): 1-24. *(Highlighted Paper for JSM 2022, Innovations in Bayesian Learning Session).*
- **Yinpu Li**, Siqi Mao, Yaping Yuan, Ziren Wang, Yixin Kang and Yuanxin Yao. Beyond Tides and Time: Machine Learning's Triumph in Water Quality Forecasting. *American Journal of Applied Mathematics and Statistics*. 2023; 11(3):89-97. doi: 10.12691/ajams-11-3-2
- Bradley Hupf, **Li, Yinpu**, Rachael Liu, and Jianchang Lin, "Predicting Phase III Results by Incorporating Historical Data using BART Extensions". *Book: Case Studies in Innovative Clinical Trials* (2023). Cat. No.: 517643. Taylor & Francis Group, LLC. https://doi.org/10.1201/9781003288640
- **Li, Yinpu** and Antonio R. Linero , "An Efficient Gradient Boosting Decision Tree Algorithm that Adapt to Sparsity in High-Dimensional Setting ". *Manuscript.*

PROFESSIONAL SERVICES

- **Referee / Editorial Boards:**
  - Biometrics 2022 – Present
  - American Journal of Applied Mathematics (AJAM) 2022 – Present
- **Technical Committee for AI / ML / DL Proceeding Conferences** 2022 – Present