# Lester (Peng-Jian) Yang

pzlester@gmail.com | https://lesterpjy.me | https://github.com/lesterpjy

## Experience

**Logitech** Hsinchu, Taiwan

Senior Engineering Scientist Jun 2022 - Present

Engineering Scientist May 2021 - Jun 2022

- Led data science innovation projects on the global CMF (color, material, and finish) engineering team.
- Enabled objective product color management by developing a patented computer vision system for high precision color measurements applicable for both engineering validation testing and production quality control.
- Introduced method for reducing carbon footprint of a manufacturing process by 11% by analyzing power consumption and production volume data, and explaining high-power contributing factors with Shapley additive explanations of XGBoost models.
- Hired and mentored data science interns, fostering their growth and skills, resulting in the prototyping of new software solutions. Under my mentorship, half of these innovation prototypes matured into official, operational projects, enhancing the team's technology tool stack.

**Omnistream Co.** Singapore

Data Science Intern Apr 2020 - Jun 2020

Omnistream.co is a retail optimization startup in Singapore.

- Worked with the data engineering and business teams on refactoring the data pipeline, improving product quality, and automating business analytics.
- Expedited product delivery by more than 300% by replacing manual estimation of product feature importance with customized XGBoost package and by automating business analytics.
- Reduced the number of XGBoost model features by 91% while maintaining model accuracy through feature engineering. Increased interpretability led to higher confidence of an onboarding client.

**Nagoya University** Nagoya, Japan

Undergraduate Researcher in Theoretical and Quantum Chemistry Aug 2018 - Sep 2019

Spearheaded a research project that piloted the application of artificial neural networks for simulating quantum chemical calculations. **Published on peer-revied ACS JCTC**. Key words: ab initio calculations, Metropolis-Hasting algorithm, variational methods, Markov chain Monte Carlo, Boltzmann machines, KL divergence

## Patents and Publications

- Peng-Jian Yang, Shang-Yu Yu, Yi-Hsieh Wang, Ko Chun Lin, Zhaoyan Han. 2023. Color matching system for product manufacturing. USPTO. US11847799B2.
- Peng-Jian Yang, Ying Ting Chen, Yuechan Chen, Daniel Cer. 2021. NT5?! Training T5 to Perform Numerical Reasoning. *arXiv* preprint *arXiv: 2104.07307v2*
- Peng-Jian Yang, Mahito Sugiyama, Koji Tsuda, Takeshi Yanai. 2020. Artificial neural networks applied as molecular wave function solvers. Journal of Chemical Theory and Computation. *doi:10.1021/acs.jctc.9b01132*

## Open Source Contributions

**LangChain**: integration of PDF document loader via pdfplumber, thus enabling usage of pdfplumber's visual debugger with LangChain

## Education

University of California, Berkeley

**Master of Information and Data Science** August 2019 – May 2021

Relevant courseworks: Statistics for Data Science, Applied Machine Learning, Natural Language Processing with Deep Learning, Experiments and Causal Inference

Nagoya University

**Bachelor of Science in Chemistry** October 2015 – September 2019

## Skills

**Statistics**: linear models, regression, hypothesis testing; **Python**: NumPy, Pandas, Matplotlib, Plotly, scikit-learn, OpenCV, PyTorch; SQL, Git, LaTex

## Projects (more on lesterpjy.me)

- **Influencing vaccination opinions**: A difference-in-difference analysis of random experiment data studying the effects of exposing engaging media of COVID-19 vaccination facts on inclination of vaccination.
- **Analysis of BayWheels Bike-share Data**: Queried top travel routes separated by user type by querying the public bike-share dataset on BigQuery with SQL. Devised recommendations for improving revenue stream with bike station status, popularity of stations by usage, and time of use.
- **Flight Delay Prediction**: prediction of departure flight delay using the Bureau of Transport Statistic on-time performance dataset, and weather data provided by NOAA. 11 gigabytes of data were cleaned, explored, and engineered with PySpark to build a gradient boosted tree model that predicts departure delay with a recall of 86%

## Teaching Experience

**Exploratory Data Analysis and Dashboarding with R** March 2023

Teaching Assistant, with Dr. Gunnar Kleemann at UC Berkeley, for US Centers for Disease Control and Prevention

## Distinctions

NGK Insulators Scholarship 2017-2019, Daiko Foundation Scholarship 2017, Japan Gov. Monbukagakusho Honors Scholarship 2016.

## Natural Language

Native Mandarin Chinese | Native English | Fluent Japanese