Fangyuan Li

(510) 926-5914 | fangyuan_li@berkeley.edu | Berkeley, CA | GitHub

EDUCATION

University of California, Berkeley — **Berkeley, CA**
M.A. in Statistics — Expected Dec 2024
- Courses: Introduction to Machine Learning, Statistical Computing, Introduction to Statistics at an Advanced Level

Zhejiang University — **Zhejiang, China**
B.S. in Statistics, Data Science Track — June 2023
- GPA: 3.87 / 4.0 (Last two years GPA: 3.93 / 4.0)

TECHNICAL SKILLS

- **Programming**: Python (Pandas, Sklearn, PyTorch), R, C, Linux, MATLAB, JavaScript, LaTeX
- **Database & Data Processing**: SQL, Dask, DolphinDB
- **Machine Learning**: Supervised (XGBoost, Neural Net, SVM, etc), Unsupervised (K-means, PCA, etc)
- **Visualization**: Python (Matplotlib, Seaborn), R(ggplot2), Excel, PowerPoint
- **Others**: Git, Web Scrapping, Docker, A/B Testing

PROFESSIONAL EXPERIENCE

DolphinDB, Inc. — **Zhejiang, China**
Data Science Intern, awarded as excellent intern — July 2022 – December 2022
- Optimized the computing efficiency by over **200** times by replacing the naïve iterative computing method for Net Asset Value Index with **parallel matrix computation** through internal C++ libraries
- Realized the panel calculation and the streaming calculation and **Newton's Method** for Guotai Alpha 191 factors using DolphinDB (Python + SQL), which was published as a new module in DophinDB and attracted **1100+** customers to watch the live feature launch
- Researched the applications of pivot table in several industry scenarios, conducted **data analysis**, and benchmarked the pivoting performance of DolphinDB against Python; Published the research deliverable in the product manual
- Built **XGBoost**, **MLP** (Multi-layer Perceptron), and **RNN** (Recurrent Neural Network) to perform real-time demand forecasting for the bike-sharing system; Achieved 0.0534 RMLSE through RNN

xQuant Technology Co.,Ltd — **Zhejiang, China**
Quantitative Analyst Intern — April 2022 – June 2022
- Optimized the investment portfolios for clients by implementing eigenvalue adjustment to the factor risk matrices in the Barra model, ran 3,000 **Monte Carlo simulations** in Python, and reduced the bias by **20%**
- Derived the formula of ARC (Active Risk Contribution) using **multivariate statistical analysis**, helping the team better understand the statistical principles of the multi-factor model

PROJECT EXPERIENCE

Analysis of the Volatility of China Financial Markets Based on the View of Hurst Statistics — Apr 2022 – Jun 2022
National Statistical Modeling Competition for College Students in China
- Conducted literature review and used the filter method and the **pre-averaging approach** to mitigate microstructure noise in high-frequency financial data when estimating Hurst Exponent of the integrated volatility
- Implemented the algorithm to calculate Hurst Parameter using MATLAB, examined the estimation using simulated financial data by implementing **cross-validation**, random seeds, and method of weighted mean to select parameters
- Incorporated the realized volatility of the China stock market into this aforementioned algorithm to estimate the Hurst Exponent and conducted a **statistical hypothesis test** to ensure the reliability of the results
- Submitted a paper *Estimation of Hurst Parameter of China Financial Market* and won the provincial first prize

Solar Panel Energy Prediction Study Using Machine Learning (GitHub) — Jan 2022 – Fab 2022
- Leveraged **Time Series Analysis** to predict the energy consumption of solar panels, built and refined **MLR**, **SVR**, and **NN** models to predict heat load; Achieved the lowest RMSE of 1.047 with SVR
- Implemented **K-means** to cluster weather conditions in the U.S. and used ggplot2 to visualize the clustering result, which helped the process of decision-making of whether to install the solar panel
- Summarized the research in a poster, delivered the presentation, and won the Honorable Mention

'Webwork' Online Homework System Log Analysis — Apr 2021 – May 2022
- Cleaned the log data collected from the online system, conducted **correlation analysis** on students' scores and their homework behaviors, and detected/removed outliers through Q-Q plot
- Implemented a **Genetic Algorithm** in MATLAB using relevant features of the questions as the fitness factors, designed and built an Automatic Test Paper Generation System which generated problem sets based on teachers' needs
- Improved the functionality of the system, which had been adopted by math teachers throughout Zhejiang University