Qin Jiayue

Singapore | qinjiayuekkk@163.com | 65 8084 4215 | linkedin.com/in/jiayue-qin-data | github.com/qjy

Education

National University of Singapore, MS in Data Science and Machine Learning	Aug 2024-Dec 2025
East China Normal University, BS in Statistics	Sep 2020-Jun 2024

- **Key Courses:** Probability Theory and Mathematical Statistics (98), Machine Learning (99), Time Series Analysis (92), Bayesian Statistics(100), Biostatistics(95), Statistical and Computational Methods in Biomedicine(96)
- **Awards:**University-level Scholarship (Second and Third Prize)

Internships

Data Modeling Intern, LVMH – Shanghai, CH   Feb 2024 – Jul 2024

- Transformed about 100,000 users datasets into user features using SQL and built an XGBoost classification model to predict customer purchasing behavior, achieving a high accuracy (AUC of 0.8) and increasing purchase rates by 350% to 440%.
- Developed dashboards to monitor model performance (AUC, F1 score, etc.), purchase rate and data source anomalies in order to evaluate and optimize models by using QuickBI.
- Conducted feature quality checks and implemented Python scripts for automation, saving time.

Research Intern, Zhongyan Technology – Shanghai, CH   Jul 2023 – Nov 2023

- Visualized survey data through bar charts, word clouds, and other methods to analyze feedback from McDonald's events, presenting enhancements in reports to improve customer experience.
- Executed significance tests on satisfaction metrics across different demographics and regions, aiding in product positioning.
- Used web scraping techniques to extract store information and implemented automated processes to extract and analyze review keywords.

Projects

Prediction of Chinese NEV Sales and Evaluation of Regional Policy Effects   Jan 2024 – May 2024

- Collected data from macroeconomic, policy and product perspectives using sources like WIND and government websites; cleaned the data using NumPy and Pandas, resulting in 11 features.
- Built SARIMA and LSTM models to predict NEV sales, with the final multivariate LSTM model achieving a higher accuracy of 12.91% MAPE.
- Utilized the Synthetic Control Method to demonstrate that two non-subsidy policies increased NEV sales in China.

Prediction of Elderly Health Status Using Stacking (Kaggle Top 15%)   May 2023 – Jun 2023

- Visualized data distribution using violin plots, heatmaps and other methods.
- Compared models using balanced log loss and 5-fold cross-validation. Optimized parameters for CatBoost and LightGBM with Optuna. Evaluated variable importance using Gini and SHAP. Built a stacking model, achieving a balanced log loss of 0.405.

Breast Cancer Dataset Analysis Based on Bayesian Methods and Frequentist   May 2023 – Jul 2023

- Employed three distinct parameter estimation methods during the construction of the logistic regression model: the HMC algorithm, MLE and Gibbs sampling.
- The Frequentist Logistic Regression exhibited the best overall performance, considering accuracy and runtime.

Skills and Languages

Programming Skills: Python, SQL, R, QuickBI, SPSS, ThinkCell
Analytical Skills: Machine Learning, Experimental Design, Data Visualization, Data Crawling
Languages: Proficient in English and Mandarin   **Hobby:** Guzheng Level 10