Qin Jiayue

Singapore qinjiayuekkk@163.com 65 80844215 Qinjiayue qjymary

Education

National University of Singapore *Aug 2024 – Dec 2025*
MS in Data Science and Machine Learning

East China Normal University *Aug 2024 – Dec 2025*
BS in Statistics

- GPA: TOP 20%
- **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)

Skills and languages

Programming Skills: Python(pandas, numpy, sklearn, keras), SQL, R, QuickBI, SPSS, Thickcell

Analytical Skills: Machine Learning, Experimental Design, Data Visualization, Data Crawling

Languages: Fluent in English and Mandarin (both spoken and written) **Hobby:** Guzheng level 10

Internships

Data Modeling Intern *Shanghai, CH*
LVMH *Feb 2024-Jul 2024*

- Transformed large data 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%-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, implemented Python scripts for automation, saving time.

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, government websites; cleaned 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 (12.91% MAPE).
- Utilized 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, etc.
- Compared models using balanced log loss and 5-fold cross-validation. Optimized parameters for CatBoost , LightGBM with Optuna. Evaluated variable importance with Gini and SHAP. Built a Stacking model, achieving a balanced log loss of 0.405.

Integration of GWAS Data for Related Diseases *May 2023-Jul 2023*

- Conducted integrative analysis of bipolar disorder and schizophrenia datasets using the EM algorithm, identifying more disease-related SNPs than single-disease analysis.
- Improved the accuracy of variant detection through multi-disease analysis.
- Explored the GPA model to enhance identification of risk variants.