李永濬 Rain Li

Master of Science in Statistics

zx20628470@gmail.com

0965823239

Linkedin

Github

EDUCATION

- **National Chengchi University**
  Master of Statistics
  2023/9~2025/7

- **Fu Jen Catholic University**
  Bachelor of Statistics and Information Science
  2018/9~2022/6

CERTIFICATIONS

- DP-100 Azure資料科學家
- TQC+ 程式語言Python 3 / 網頁資料擷取與分析Python 3
- Google Analytics (分析) 個人認證

SKILLS

- *Data & Analytics:*

  SQL (BigQuery), ETL Pipelines, EDA, Statistical Analysis, Experiment Design

- *Machine Learning & AI:*

  Statistical Modeling, Deep Learning (PyTorch, TensorFlow), ML Algorithms, Embeddings, RAG, Retrieval Systems, Tool Calling, MCP, A2A

- *Engineering & Tools:*

  Python, RESTful API Development, Docker, Git/GitHub, Kubernetes (Basic), Power BI

- *Methodologies: Agile/Scrum*

PROFILE — *Data Scientist (AI / Machine Learning Focus)*

- **Expertise in Modeling & Modern AI: Specialized in statistical learning and Bayesian modeling, with hands-on experience in RAG architectures, LLM fine-tuning, and building Python-based ML pipelines for large-scale data.**

- **Solid Theoretical Foundation: Advanced training in deep learning and statistical inference through graduate research at NCCU, enabling rigorous approach to complex data challenges.**

- **Business-Tech Liaison: Strong capability in integrating AI APIs and translating analytical results into deployable services and actionable insights for business and product teams.**

Work Experience

2024/11 – 2025/2

類神經網路股份有限公司Aiii.Ai

Data team intern

- Conducted large-scale data analysis using Google BigQuery and SQL
- Familiar with API development and integration
- Exposure to advanced AI topics: RAG, fine-tuning, vector databases

2024/7 – 2024/8

台北富邦商業銀行股份有限公司 Fubon

數位科技研發部/雲端發展應用科實習生

- Engaged in cloud technology research and application
- Assisted in the development, testing, and maintenance of cloud-based systems

2023/6 – 2023/11

緯創軟體 WITS

Trainee : Digital Health & Analytics Accelerator Program

- Learned and applied cloud services, SQL, and Python data analysis
- Integrated and analyzed data using ChatGPT API
- Created data reports with Power BI

AWARD

- ***First place*** *in the 2021 Inno Serve Business Information Innovation and Applications category.*

PROJECTS

- ## Master's Thesis: Sports Analytics with Bayesian Skill Updates and Deep Neural Interaction Models

Designed a deep learning framework integrating Bayesian inference, feature interactions, and temporal modeling to predict sports match outcomes. Improved robustness using a frozen backbone strategy, achieving superior performance over traditional models.

- ## 食餓點伴

"食餓點伴" is a dietary recommendation and tracking app that utilizes Taipei City's ingredient information platform to organize food-related data efficiently. It employs word segmentation analysis and Jaccard similarity calculation to filter out less significant ingredients, retaining only pertinent variables for K-means clustering. The app further enhances user experience by integrating filters for allergens and diabetes-sensitive ingredients during initial registration, ensuring personalized food recommendations aligned with dietary restrictions and preferences.

- ## DIABEEATS 食在制糖

An Line Bot account that helps diabetic patients with blood glucose control. The chatbot stores blood glucose information in a cloud database, allowing diabetics to view changes in blood glucose in real time. At the same time, we also provide personalized dietary advice to reduce the burden of diabetic patients by calculating the recommended insulin dosage based on the nutrient information of the meal.

- ## MULTI-AGENT DATA FRAMEWORK | PERSONAL PORTFOLIO

Architected a Hub-and-Spoke multi-agent system using MCP to automate SQL and Data Science workflows. Optimized local performance via 4-bit quantized Qwen-VL-Instruct and secured execution within a Python sandbox. Integrated a shared state store for persistent context, significantly enhancing system modularity, scalability, and data privacy.

- ## KKCOMPANY MUSIC CHALLENGE: SONG RECOMMENDATIONS EMPOWERED BY LANGUAGE AND RETRIEVAL MODELS

I focused on leveraging the pyserini library for retrieval and integrating KNN for handling missing values in our project aimed at song recommendation. Initially, Principal Component Analysis (PCA) was employed to select essential variables, enhancing system performance. Furthermore, I integrated the BM25 implementation from pyserini and utilized it to create text vectors, optimizing similarity calculation and feature extraction. By incorporating cosine similarity for measuring song similarity and employing KNN interpolation to handle missing values, we significantly bolstered the reliability of our recommendation system. Crucially, this project underscored the importance of teamwork, effective communication, and coordination.

PROJECT SHOWCASE