Zian Xu

6465059056 | zx2514@columbia.edu | anniexuzian.github.io

EDUCATION

Columbia University — New York, NY
M.S. in Data Science — Exp Dec 2025
- Courses: Applied Machine Learning, Database Management Systems, Algorithms for Data Science

The Hong Kong University of Science and Technology (HKUST) — Hong Kong
B.S. in Data Science and Technology, **First Class Honors**, *GPA: 3.96/4.3(Top 1%)* — July 2024
- Courses: Object-Oriented Programming & Data Structures, Algorithms, Machine Learning, Computer Networks
- Honors: Academic Achievement Medal (Top 1% of Graduating Class), HKSAR Scholarship, Dean's List

SKILLS

- **Programming Languages:** Python, Java, C++/C#, JavaScript, SQL, R, HTML5/CSS, TypeScript
- **Frameworks**: Flask, FastAPI, Node.js, Express.js Vue.js, React, AngularJS, Django
- **Data & Machine Learning Tools:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, TensorFlow, PyTorch
- **Others:** MySQL/PostgreSQL, SQLite, AWS/GCC, Git/GitHub, Docker, Software Engineering (Frontend, Backend)

PROFESSIONAL EXPERIENCE

TAL Education Group — Beijing, China
Machine Learning Engineer Intern — May 2023 - Aug 2023
- Developed an automated data pipeline in Python for processing over 20,000 HR records, incorporating robust data cleansing and transformation practices that improved data quality and consistency for machine learning models.
- Built and optimized machine learning models, including Neural Networks, XGBoost, Random Forest, to predict employee commitment, resulting in a 15% reduction in mean absolute percentage error through cross-validation.
- Analyzed model results and identified key factors using feature importance; delivered insights through Tableau dashboards to leaders, improving decision-making processes for employee evaluation.
- Fine-tuned a reward model for Generative AI (MathGPT) to solve math problems using question-answer pairs, with a focus on hyperparameter tuning and architectural adjustments, achieving a 2% increment on the re-rank metrics.

CSE Department, HKUST — Hong Kong
Teaching Assistant — Feb 2022 - May 2022
- Hosted lab sessions for 90 students in *COMP2011-Programming with C++*, elaborating tasks, answering questions, and facilitating demos and grading logistics

RESEARCH EXPERIENCE

Course Recommendation System Development — June 2023 - May 2024
- Led a team of 4 to develop a personalized course recommendation platform using a Graph Neural Network (GNN).
- Trained and fine-tuned a heterogeneous graph neural network (HetGNN) for course grade prediction using PyTorch, improving the model prediction accuracy by 20%.
- Developed a user interface using Vue.js, incorporating responsive layouts and visually engaging elements to enhance user experience and accessibility.
- Engineered and optimized backend REST APIs using Express.js to interact with the HetGNN model, MySQL database, enabling high-performance query handling and efficient recommendation computations.

Language Model Evaluation by Natural Language Processing — May 2023 - Aug 2023
- Developed a nine-dataset evaluation framework to assess language model performance in the financial domain.
- Compared the performance of encoder-only language models vs. decoder-only language models (BERT vs. GPT).
- Investigated the efficacy of in-context learning for ChatGPT with different sample selection strategies, finding that in-context learning is more effective for complex tasks.

Interactive Data Storytelling in VR — May 2022 - Sept 2022
- Developed an interactive VR story and user interface in Unity using C#, improving user situation awareness of health risks by 50% compared with traditional desktop presentations.
- Conducted iterative testing with 66 participants, using A/B testing and data visualization in Matplotlib and Pandas to drive design improvements.