Zian Xu

(646) 505-9056 | zx2514@columbia.edu | anniexuzian.github.io

EDUCATION

Columbia University New York, NY
M.S. in Data Science, GPA: 4.17/4.0 Exp Dec 2025
- Courses: Applied Machine Learning, Exploratory Data Analysis & Data Visualization, 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* July 2024
- Courses: OOP& Data Structures, Machine Learning, Statistics & Probability, Algorithms, Time Series, Databases
- Honors: Academic Achievement Medal (Top 1% of Graduating Class), HKSAR Scholarship, Dean's List

PROFESSIONAL EXPERIENCE

TAL Education Group Beijing, China
Machine Learning 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), optimizing hyperparameters and architecture to improve mathematical problem-solving accuracy, resulting in a 2% improvement on re-rank metrics.

CSE Department at 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).
- Built an ETL pipeline with Python to process over 40K student-course records efficiently, using SQL for data extraction and transformation.
- Trained and fine-tuned a heterogeneous graph neural network (HetGNN) for course grade prediction using PyTorch, improving the model prediction accuracy by 20%.
- Developed various backend APIs to interact with the HetGNN model, MySQL database and perform complex logic.

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 and decoder-only language models, such as Bert and GPT, on financial NLP tasks and concluded the pros and cons of different models.
- 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
- Implemented 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 three rounds of user studies with 66 participants, employing A/B testing to assess VR interaction effectiveness, gathering valuable feedback for further improvements.
- Summarized top design elements by analyzing and visualizing experiment data using Matplotlib, Pandas and NumPy.

SKILLS

- **Programming Languages:** Python, Java, C++/C#, JavaScript, SQL, R, Matlab, HTML5/CSS
- **Data & Machine Learning Tools:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, TensorFlow/PyTorch/Keras
- **Others:** MySQL/PostgreSQL, SQLite, AWS/GCP, Git/GitHub, BeautifulSoup/Selenium, PowerBI/Tableau, Spark