JU, Long (Logan)

Linkedin | loganju@umich.edu | 734-754-0635 | Seeking Data Science& Machine Learning Internship (2025 summer)

EDUCATION

University of Michigan, Ann Arbor Ann Arbor, MI

Master of Science in Electrical and Computer Engineering | Track: Machine Learning Aug 2024- Present

- Coursework: Natural Language Processing, Database Management System, Data Structures and Algorithms

University of California, Berkeley Berkeley, CA

Visiting Student at EECS | GPA: 3.95/4.0 Jan 2023 – Dec 2023

- Coursework: CS189 Machine Learning, CS182 Deep Learning, DataC100 Tech & Principal of Data Science, CS61A Structures of Programs, EECS127 Optimization, STAT151A Linear Modelling

Nanjing University Jiangsu, China

Bachelor of Science in Statistics | GPA: 4.2/5 Sep 2020- Jun 2024

- Coursework: Probability, Linear Algebra, Mathematical Statistics, Data Analysis, Numerical Analysis
- Honors: Nanjing University Honor Program for Mathematics (2020-2021, 2021-2022)

SKILLS

- **Programming:** Python (PyTorch, pandas, numpy, scikit-learn, seaborn, matplotlib), SQL, R, C++, MATLAB, LaTeX
- **Technical Skills:** Machine Learning, Deep Learning, Natural Language Processing, Data Wrangling & Feature Engineering, Statistics, Hypothesis Testing, A/B testing, Time Series Analysis (ARIMA, SARIMA, LSTM)
- **Tools:** AWS, Azure, LangChain, LlmaIndex, Huggingface, Git, Tableau, Power BI, APIs

WORK EXPERIENCES

Coursera & Berkeley DS Discovery Program Berkeley, CA

Data Analyst Sep 2023 – Dec 2023

- Optimized the data pipelines by leveraging **SQL and PySpark** for efficient data integration, cleaning, and transformation of large datasets (millions of rows) from Coursera's Product Sales data (from 2018-2023).
- Developed interactive visualizations using **Tableau and Matplotlib** to present key insights from previous A/B tests to the management team, enabling data-driven decision-making.
- Applied **Random Forest** Regression to analyze the relationship between exchange rate fluctuations and revenue, optimizing pricing strategies in Asian countries which resulted in a 10% increase in revenue from new users.
- Utilized **Gradient Boosting** to analyze price relationships across different products, determining the optimal prices in North America resulting in a 12% increase in the number of new users.

SELECTED PROJECTS

Rag-Based LLM with Memories May 2024-Aug 2024

AI Engineer at Skylow, Berkeley Entrepreneur Association

- Developed the AI characters API utilizing **Retrieval-Augmented Generation (RAG)** technology with a memory layer, enabling continuous learning from user interactions.
- Designed and implemented a hierarchical structure for data embedded in the vector store, enhancing the organization of chunks and improving retrieval effectiveness.
- Enhanced RAG technology by integrating **multi-step query transformations, Hyde, and step-back prompt methods**; deployed a **re-ranking system** to re-evaluate retrieved chunks, resulting in increased retrieval accuracy.

AI-Driven Street View Modifications in Urban Planning (poster link) Oct 2023- Jan 2024

CS182 Deep Learning Project at UC Berkeley

- **Overcame Stable Diffusion's 72-token limit** by chunking long prompts, encoding each chunk separately, and concatenating the embeddings to drive the image generation process.
- Implemented **prompt-to-prompt stable diffusion** to generate the finetune dataset, ensuring the original image structure is preserved while generating new urban designs from modified prompts.
- Finetuned a **pix2pix** model aimed at transforming street views based on criteria like sustainability and aesthetics, using inputs with a starting image and specific prompts. **(model link)**

ADMM For Semi-Definite Program with Burer-Monteiro Reformulation Jun 2023 – Sep 2023

Research Assistant at Shanshu Technology Beijing Co Ltd

- Reformulated General SDP into bi-linear and convex formats to facilitate solution via ADMM algorithm, enhancing efficiency in solving complex optimization problems.
- Replaced the gradient descent methods with linear system in each iteration of the optimization process, providing **convergent** solutions on various forms of SDP.
- Converge **15%** faster than the traditional Burer-Monteiro methods on SNL, MAX-CUT, and MAX-Bisection.