Meichen Liu

Mobile: +1(780)6952676 | Email: meichen1@ualberta.ca | LinkedIn: /in/meichenliu5577/ | GitHub: meichen1

Skills

- Programming: Python, PySpark, SQL, R, JavaScript, Excel VBA, Linux
- Software and Platform: Databrick, Azure Data Factory, GCP, AWS, Docker, JIRA, GitHub, SharePoint, Power Automate, Power BI, Tableau, ChatGPT, Langchain, Hugging Face
- Machine Learning and Statistics: NumPy, SciPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Hypothesis Testing, A/B Testing, ARIMA, PCA, LDA, RF, XGBoost, CatBoost, K-means, KNN, DBSCAN, NLTK, TensorFlow, PyTorch

Education

Ph.D in Statistics — 2018.09 - 2024.04
University of Alberta.
- GPA: 4.0/4.0, Alberta Graduate Excellence Scholarship (AGES)

Experience

Senior Data Analyst — 2024.05 - Present
London Health Science Center / Western University — London, ON

ECG Research RAG Chatbot

- Developed a specialized LLM chatbot to streamline research by retrieving the latest ECG-related articles from sources like arXiv, MedPub, and Google Scholar. This project focused on empowering researchers with timely data-driven insights.
- Fine-tuned a **Llama**-based LLM and constructed a **RAG** (Retrieval-Augmented Generation) database. Utilized **Pinecone, LangChain, Streamlit, and Hugging Face** to harmonize information retrieval, enhancing the chatbot's sophistication in delivering contextually relevant academic content. Resulted in improved decision-making and accelerated learning curves.

Arrhythmia Treatment Predictor

- Led the project "AF-CARE-AI," which focuses on predicting the success rate of distinct atrial fibrillation (AF) treatments, providing cardiologists with data-driven patient management and improving overall treatment efficacy.
- Employed convolutional neural network (CNN) and **Transformer** models for predicting AF treatment outcomes. Data processing includes standardizing ECG signals and extracting features from echocardiogram images.
- Deployed this predictive tool in clinical practice, resulting in a 20% recall increase, equivalent to a patient retreatment cost reduction of ~$2 million/year.

Data Scientist / Statistical Consultant — 2018.09 - 2024.04
University of Alberta — Edmonton, AB

Multimodal Alzheimer's Disease Identifier

- To help and coordinate with the medical research team elevate the accuracy of AD diagnosis, built a multimodal binary classification system using brain imaging data obtained from ADNI [Link], a longitudinal multicenter study.
- Combined the patients' demographic and brain image features through kernel embedding; trained and optimized multiple classification models (**Logistic regression, SVM, Random Forest, XGboost, LightGBM**).
- The best SVM model enhanced AD prediction accuracy by 50% over the baseline system and the recall to 95%, potentially enabling 2M patients every year in Canada to receive early treatments. [Paper]

Fairness enhancement model of AI systems

- To detect and mitigate gender and racial bias in job postings that may violate the **AI ethics** requirements, developed a debiasing algorithm to alleviate potential unfairness, and collected data from job boards like Indeed and Monster.
- Managed the project to design unfairness metrics and trained multiple models including **Partial least squares (PLS)**, **Quantile regression** with **regularization (L1, L2)** and **Multilayer Perceptron (MLP)**.
- Helped the minority gain 20% more opportunities in job markets. Publications on top ML conferences AAAI 2022 [Paper], Neurips 2022 [Paper], and designed a BIAS project website [Link] based on Jekyll, Bootstrap 4.6.

Project

Cyberattack Detection Model — 2023.12 - 2024.03

- To help the consulting firm effectively detect cyberattacks from large amounts (3 million) of Edge-IIoT network traffic records, discover underlying factors, and provide actionable safety suggestions.
- Collaborated with the web development team to analyze user activities, engineer features, train and optimize multiple models (**Logistic Regression, Random Forest, XGBoost, CNN, and Autoencoder**), and apply SHAP for interpretability.
- Deployed the best model on Microsoft Azure as a Web App using Docker and Flask, optimizing cloud resources to achieve a precision of 0.96 and recall of 0.95, resulting in annual cost savings of $100K [GitHub].