Mingyuan Xu

Software / Data Engineering Intern (UNSW Distinction)

mingyuanxu.unsw@gmail.com +61 402 479 470 Matraville Sydney, NSW
https://github.com/MasterpieceXu

Summary

Detail-oriented Master of IT (AI) student at **UNSW** with a consistent **Distinction average**. Passionate about **Data Engineering** and **Backend Development**, with proven experience in architecting RESTful APIs, optimizing high-dimensional data pipelines, and implementing robust security protocols. Seeking a Software/Data Engineering Internship to leverage my skills in Python, Cloud systems, and end-to-end system design.

Education

The University of New South Wales (UNSW) Master of Information Technology • Distinction Average (WAM: 77.7 / 100)
Information Technology (Artificial Intelligence) Sydney, Australia • May 2025 - May 2027 (Expected)

Experience

CUG(China University of Geosciences) Research & Development Base Zigui, China
Internship 2023.6-2024.6

- **Automated Data Cleaning & Visualization Pipelines** for groundwater monitoring datasets using Python, increasing data processing efficiency for the research team.
- **Developed Random Forest Regression Models** to predict water quality indicators, utilizing **Feature Importance** analysis to identify key environmental impact factors.
- **Synthesized Technical Reports** based on model outputs and data visualizations, supporting the delivery of scientific research projects.

Projects

NSW Public Transport GTFS Data Service System 2025.06 – 2025.12

- **Architected a high-performance Data Engine** to parse and process complex GTFS (General Transit Feed Specification) datasets, designing a relational schema to optimize spatial-temporal queries.
- **Developed robust RESTful APIs** using Flask-RESTX, implementing Geospatial Indexing to support real-time location-based searches and data exports.
- **Enhanced System Security & Scalability** by integrating **JWT (JSON Web Tokens)** for stateless authentication and implementing custom middleware for request validation.
- **Containerized the application using Docker** for seamless environment consistency and **deployed via Azure App Service**, ensuring high availability and automated scaling of the data service.
- **Engineered Interactive Documentation** via Swagger UI, streamlining API integration and reducing debugging overhead through standardized response structures

https://github.com/MasterpieceXu/NSW-Transport-API

Financial Fraud Detection and Transaction Value Prediction 2025. 10 – 2025.12

- **Engineered a multi-task machine learning pipeline** to process a large-scale dataset of **1 million transaction records**, achieving high-precision fraud detection (F1-Macro: 0.91) and amount prediction (RMSE: 29.93).
- **Developed complex Geospatial & Temporal features**, including **Haversine distance** for anomalous location detection and **rolling statistics** (`cc_time_since_last`) to identify rapid, consecutive transaction patterns.
- **Implemented sophisticated data aggregation strategies**, such as category-based mean normalization (`amt_vs_cat_mean`), to establish behavioral baselines and significantly enhance model sensitivity to relative spending outliers.
- **Addressed extreme class imbalance (0.52% fraud rate)** by integrating **SMOTE oversampling** with **LightGBM**, utilizing Bayesian threshold optimization to ensure robust model generalization under strict computational constraints.
- **Optimized predictive performance** by selecting **Ridge Regression** for linear-correlated transaction amount tasks and **LightGBM** for non-linear fraud patterns, balancing system speed with memory efficiency.

https://github.com/MasterpieceXu/Financial-Fraud-Detection-and-Transaction-Value-Prediction

Reliable Data Transfer Protocol over UDP 2025. 10 – present

- **Architected a reliable transport layer protocol** on top of connectionless UDP, ensuring 100% data integrity and ordered delivery in lossy network environments.
- **Implemented core TCP-like mechanisms**, including sequence numbers, acknowledgments (ACKs), and checksums to detect and recover from packet loss, bit errors, and reordering.
- **Engineered advanced Flow Control & Congestion Control** algorithms (such as **Sliding Window** or **Go-Back-N**), optimizing throughput while preventing network congestion.
- **Developed a robust retransmission logic** with dynamic timeout estimation, significantly improving data transfer stability under varying network latency conditions.
- **Conducted rigorous stress testing** using simulated network emulators to verify protocol resilience against high packet loss rates and extreme jitter.

https://github.com/MasterpieceXu/URP-UDP-based-Reliable-Protocol

Skills

Software & Data Engineering

Python (Proficient), SQL, Flask-RESTX, SQLAlchemy, PostgreSQL, RESTful API Design, ETL Pipelines, JWT Authentication, Data Modeling

AI & Machine Learning

PyTorch, Transformers, LightGBM, XGBoost, Scikit-learn, Pandas, Feature Engineering

Systems, Cloud & DevOps

Docker (Containerization), Azure App Service, Linux/Unix, Git, Socket Programming, TCP/UDP Protocols, Multi-threading