

Hang Ruan

hangruan27@gmail.com | (551) 359-9445 | <https://www.linkedin.com/in/hang-ruan-963306205/>

EDUCATION

Carnegie Mellon University

Master of Information Systems Management, Business Intelligence & Data Analytics

August 2024 - December 2025

Core Courses: Intro to Deep Learning / Machine Learning with Large Datasets / Database Management / Business Analytics / Data Analytics / Time Series Forecasting in Python / Object Oriented Programming in Java / A/B Testing / Decision Models

New York University

September 2020 - May 2024

Bachelor of Arts, Economics & Data Science

Core Courses: Advanced Econometrics / Data Structures and Algorithms / Machine Learning / Multivariable Calculus / Linear Algebra / Linear and Nonlinear Optimization / Money and Banking / Competitive Analysis / Market Design / Game Theory

SKILLS

- **Programming & Tools:** Python (pandas, matplotlib, spacy, Beautiful Soup, sklearn, pytorch), Java, Stata, R, Tableau, Power BI
- **DS & ML:** Machine Learning (KNN, SVM, GBDT), Deep Learning (CNN, RNN, GNN), Reinforcement learning (PPO, A2C)
- **Database & Big Data:** MySQL, MongoDB, Google BigQuery, Hadoop MapReduce, Apache Spark, AWS, GCP, Databricks

PROFESSIONAL EXPERIENCE

Spotify

New York, United States

Data Analyst Intern

June 2025 - August 2025

- Constructed a standardized dataset of 1,500+ climate events and migration detection algorithms with Google BigQuery to measure gross vs. net impacts on core business metrics including **streams, DAU, listening hours, and ad revenue**
- Developed a historical impact analysis pipeline in Python to uncover patterns by event type, severity, geography, and seasonality, and built **Linear Regression, ETS, and Random Forest models** to forecast disaster impacts and recovery timelines
- Generated long-term business impact projections (2030, 2050, 2100) across multiple climate scenarios, identifying ~\$60M in global revenue at risk, and presented climate risk heat maps through **interactive dashboards in Google Looker Studio**

Deloitte

Shanghai, China

Data Science Intern

February 2024 - May 2024

- Designed an **automated data pipeline** to extract data via APIs, web scraping, and LLM-based report parsing, **clean and normalize** more than 10,000 raw inputs, and **load structured datasets** into databases for downstream tasks
- Identified key feature factors using **PCA** and **correlation analysis**, and applied **data discretization techniques** such as **Chi-merge, decision tree binning, and LOWESS smoothing** to enhance model robustness and mitigate overfitting
- Deployed **Logistic Regression** and **XGBoost** models to predict default probabilities and generate credit ratings for 500+ firms, achieving 0.85 F1 Score and 0.95 SMD, enabling clients to assess default risk and make informed investment decisions

China Merchants Securities

Shanghai, China

Capital Market Intern

September 2023 - November 2023

- Supported deal structuring, pricing report drafting, and pitch deck preparation, and leveraged **Tableau** to visualize real-time credit spread movements, contributing to \$2B in corporate debt offerings and reduce financing costs by ~15%
- Built **ARIMA and LSTM models** for short-term interest rates prediction, and implemented **moving-average smoothing** to reduce data noise, achieving forecast error under 30 bps and informing issuance timing and pricing decisions

PROJECT EXPERIENCE

Video-Text Retrieval with CLIP4Clip

January 2025 - May 2025

- Fine-tuned CLIP4Clip model with four different similarity modules (**Min Pooling, Max Pooling, LSTM, Transformer**) to strengthen multi-modal representation learning, achieving 39.8% R@1 (text-to-video) and 40.8% R@1 (video-to-text)
- Implemented three different frame-selection methods (**scene-change, hybrid allocation, uniform sampling**) to optimize visual-text alignment and reduce temporal redundancy, reaching 39.6% R@1 (text-to-video) and 40.5% R@1 (video-to-text)

Pruning-Based Efficient Model Compression

September 2025 - December 2025

- Ran controlled experiments across nine pruning methods, including **magnitude-based pruning, L1-norm filter pruning, ThiNet, DepGraph, and WANDA**, to evaluate accuracy-efficiency trade-offs in neural network models
- Developed an end-to-end training framework with **data preprocessing, pre-training, pruning, fine-tuning, and evaluation**, achieving up to ~97% sparsity while maintaining ~67% accuracy on a five-class image classification task