

CHENGJUN (RAINA) ZHANG

Chicago, IL 60601 | (740)-281-8150 | raina23278@hotmail.com | [GitHub](#)

EDUCATION

Indiana University, Kelley School of Business

Ph.D. in Business of Information Systems

August 2024 - May 2029 (expected)

Bloomington, IN

Johns Hopkins University, Whiting School of Engineering

M.S. in Security Informatics of Computer Science

August 2022 - December 2023

Maryland, MD

Denison University

B.A. in Computer Science, B.A. in Data Analytics (Economics Concentration)

August 2018 - May 2022

Granville, OH

Academic Courses: Macroeconomics, Microeconomics, Accounting, Advanced Econometrics, Applied Statistics, Data Structure, Data System, Algorithms, Machine Learning, Vision of Bayesian Inference, Natural Language Processing, Self-Supervised Learning, Security Analytic

RESEARCH EXPERIENCE

Research Assistant

Advisor Dr. Sudip Gupta – Carey Business School, Johns Hopkins University

October 2022 – Present

Baltimore, MD

❖ **Implementation of Autoencoder and Timeseries based Transformer in stock return prediction.**

- Collected, cleaned, and integrated monthly stock returns data of 96 companies from 1957 to 2016, along with 94 key stock-level predictive characteristics.
- Utilized **autoencoder** for reducing the dimensionality of returns, enhancing adaptability to nonlinear conditions and consideration of latent factors.
- Leveraged **timeseries-based transformer** model to predict stock returns using reduced dimensionality data with the predictive characteristics.

❖ **Implementation of generative model alleviates health equality using health care multimodal data.**

- Processed and vectorized 10Gb+ clinical diagnosis notes/ images, and personal information.
- Leveraged **CNN** and traditional ML models to predict the quality of drug prescribed per day.
- Leveraged **Variational Autoencoder** to generate counterfactual distributions for mitigating fairness bias.

❖ **LLMs in central bank policy changing prediction.**

- Utilized Selenium and XPath to develop a web scraping script responsible for capturing meeting minutes published by the Reserve Bank of India (RBI) spanning the years 1926 to 2023.
- Conducted **sentiment analysis** on statement text using Loughran and McDonald Sentiment Word Lists, and visualized sentiment trends with financial data over time.
- Trained **BERT** models on financial texts for predicting the RBI decisions on interest rates.

Research Assistant

Advisor Dr. Xiangyang Li - Whiting Engineering School, Johns Hopkins University

September 2022 – January 2023

Baltimore, MD

❖ **Adversarial ML attacks on spam detection system**

- Vectorized emails through **TF-IDF, D2V, and W2V**, and built a ML-based spam classifier.
- Applied **PGD** attack to manipulate vector positions in space, enhancing the spam detection system's false negative rate, allowing crafted spam emails to evade spam classifiers.
- Utilized **cosine similarity** to pinpoint "top good words" post-PGD attack with precision and efficiency.
- Achieved an 8% increase in the success rate of the cosine similarity method compared to vector aggregation methods.

PROJECTS

Innovations on vision-language pre-trained (VLP) model

- Stored and processed 14M+ images with LMDB.
- Built a VLP model with a shortcut re-weighting technique to improve the ability of the Vision Model through PyTorch.

CHENGJUN (RAINA) ZHANG

Chicago, IL 60601 | (740)-281-8150 | raina23278@hotmail.com | [GitHub](#)

- Integrated **VIT** and **BERT** and trained with masked salient tokens to maximize the utilization of computation, executing five impactful model experiments with varied structures.

Encrypted mobile devices' network traffic recognition of application type and VPN usage.

- Simulate actual manipulations by Selenium in Python for six different applications to generate network traffic data.
- Converted ten gigabytes of network packets from pcap to JSON, processed and stored them in MongoDB.
- Built a classifier utilizing **BiLSTM+Attention** model for temporal and **1D-CNN** model for spatial feature extraction.

Autonomic cyber security system on Smart Power Grid

- Simulated Smart Power Grid structure and communication between power grid and console via DNP3 protocol.
- Developed threat models and performed attack scenarios on the simulated Smart Power Grid testbed.
- Employed **GWO** algorithm for feature selection and constructed machine learning models for anomaly detection.
- Integrated the model into the ICS and implemented alarm and analysis functions to increase system resistance.

Annual Household income and household-based alcohol expenditure

- Utilized Tobit regression to verify the positive relationship between household income and alcohol expenditures.
- Create split models based on different income groups to conduct in-depth analysis for each group.

WORK EXPERIENCE

Data Scientist Intern

March 2021 – August 2021

YunChang Technology Co., Ltd (NebulaJoy)

Beijing, China

- Processed 1TB of user data with 20+ features from five game servers using MongoDB and Python.
- Utilized Tableau for data visualization, offering actionable insights to guide project decision-making.
- Employed **MATE** models to predict users' next location, achieving 61% accuracy and cutting 15% loading time.

Data Development Intern (rotational program)

March 2020 – March 2021

Wuhan Zhongyu World Technology Co., Ltd

Wuhan, China

Data Engineering Team

- Deployed ETL pipelines for 500Gb+ data, improving efficiency by 40% using **Python** and **MySQL**.
- Automated data queries and reports with **Django**, cutting down Product team demands by 80%
- Trained an in-house **XG Boost games classifier** using NLP, reducing manual labeling by 75%.

Data Science Team (Marketing and Advertising)

- Applied TF-IDF and LDA for ads' keyword extraction for tracking the latest trends of mobile game industry.
- Leveraged **A/B Testing** on ad variations to find the most suitable ads and tag words, reducing the CPA up to 38%.

Data Science Team (Products)

- Implemented **Kmeans** for user in-game behavior analysis, providing the insights on ads targeting and version adjustment; Helped improve the **conversion rate** and **ROI** of three products.
- Designed **A/B Testing** to optimize game play, boosting Day-1 retention by 23% and average ad views per user by 16%.

TEACHING EXPERIENCE

Johns Hopkins University – Teaching Assistant

October 2022 – May 2024

Investment, Empirical Finance, Linear Econometrics

Denison University – Teaching Assistant

January 2022 – May 2023

Cultural Analysis