

## SKILLS

**Data Science:** Statistical Analysis, Machine Learning, Graph Analysis, Deep Learning, Scikit-learn, TensorFlow, PyTorch

**Data Analysis & Visualization:** Python (Pandas, Matplotlib), SQL, R, Looker Studio, Jupyter

**Data Engineering:** PySpark, Airflow, GCP (BigQuery, GCS, DataProc, BigTable), Git

---

## EXPERIENCE

**Research Assistant @ Simon Fraser University** - BC, Canada

*September 2024 - Present*

- Investigating privacy-preserving techniques for network embeddings to allow data owners to publicly share user embeddings (e.g., for Data Science competitions) without compromising sensitive user information
- Implemented an edge-privacy-preserving method using Pytorch

**Teaching Assistant @ Simon Fraser University** - BC, Canada

*September 2024 - Present*

- Mentoring students from various backgrounds to understand statistics and analyze data in R

**Machine Learning Engineer @ Momo** - Top 1 fintech company in Vietnam - Vietnam

*August 2021 - June 2023*

- Reduced the fraud rate by 50% by tripling the performance of the account takeover detection model
- Collaborated with business teams to define metrics, run A/B tests, and perform analytics on 30 million users
- Developed 5+ advanced Looker Studio dashboards for monitoring OKR metrics and detecting model shifts
- Built robust data pipelines processing over 1TB of data daily with SQL, PySpark and optimized querying cost by 80%
- Scraped phone specs data and inferred missing prices to support clustering users based on price sensitivity

**Undergraduate Research Assistant @ University of Toledo** - OH, USA

*October 2020 - Dec 2021*

- Analyzed network structures of interstate disputes and presented the results at Complex Networks 2021
- Developed DyNetworkX, an open-source Python package that handles dynamic network data, and co-hosted a tutorial session for it at SBP-BRiMS 2021

**Dev-Ops Intern @ Crown Equipment Corporation** - OH, USA

*January 2020 - August 2020*

- Improved productivity of the development team by maintaining CI/CD pipelines and managing AWS resources.
- 

## EDUCATION

**Simon Fraser University** - Canada

*September 2024 - Present*

Master of Science in Statistics (GPA: 4.04/4.33)

Topics covered: Generalized Linear Models, Experimental Design, Statistical Computing, Statistical Theory

**University of Toledo** - United States

*August 2018 - December 2021*

Bachelor of Science in Data Science with Honors, Economics Minor (GPA: 3.98/4.00)

Awards: Summa Cum Laude, President's List, International Student Scholarship

Publication: Do, H. N., & Xu, K. S. (2022). Analyzing escalations in militarized interstate disputes using motifs in temporal networks. *Complex Networks and Their Applications X*, 527–538

---

## PERSONAL PROJECTS

**Credit Card Default Prediction**

*October 2020 - December 2020*

- Led and mentored 5 other Data Science club members to work on the project
- Achieved the final model with an f05 score of 0.8 by using Logistic Regression, handling unbalanced datasets with SMOTE, and securing a monotonicity relationship between predictors and target with WOE transformations