

Sinclair Schuetze

soschuetze@gmail.com | soschuetze.github.io | github.com/soschuetze

Education

University of Oxford – MSc in Social Data Science

Aug 2024

Courses: Applied Machine Learning, Network Analysis, Applied Analytical Statistics, Data Analysis at Scale

Grade: Distinction (First Class Honours)

Wellesley College – BA in Data Science and Economics

May 2023

Courses: Natural Language Processing, Machine Learning, Web Search and Mining, Multivariate Data Analysis, Statistical Inference

Awards and Honors: 3.94 GPA, Summa Cum Laude, Phi Beta Kappa, Sigma Xi

Experience

Research Fellow, Stanford RegLab – Stanford, CA

Sept 2024 – Sept 2025

- Training PyTorch Geometric GraphSAGE GNN to classify partnerships' risk of noncompliance using networks of taxpayer structures
- Implementing self-supervision tasks on heterogeneous graph such as link prediction to improve performance of GNN
- Analyze model predictions to ensure businesses identified provide greatest opportunity for increased revenue
- Utilize current ML and accounting research to guide decision making to improve the IRS's audit selection process

Data Science Intern, Mercury Insurance Group – Brea, CA

Jun 2023 – Aug 2023

- Analyzed XGBoost auto underwriting model to identify key areas for improvement, increasing predicted profit from model by 28%
- Created 15 new features using SQL queries, resulting in a 23% increase in predictive accuracy for high-risk policies
- Optimized models using SHAP and XGBoost feature importances, maintaining performance after removing 60 features
- Built R-shiny dashboard with lift charts and profit improvement visualizations, facilitating decision making by stakeholders

Software Engineering Intern, JP Morgan Chase & Co. – Wilmington, DE

Jun 2022 – Aug 2022

- Implemented contract testing framework, allowing for scalable testing of all microservice applications within data pipelines
- Developed 2 Rest APIs and updated functionality of existing APIs responsible for handling \$2 trillion in consumer payments daily
- Built a new microservice to maintain logs using Java, Spring Boot, and Kafka

Software Engineering Intern, JP Morgan Chase & Co. – Wilmington, DE

Jun 2021 – Aug 2021

- Automated 6 data pipelines using ETL framework, ingesting and transforming consumer data using Spark SQL and JPMC libraries
- Tested pipeline transformations and implemented step definitions using Cucumber files and deploying to DPL server

Projects

GNN-Driven Change-Point Detection of Trade Networks

github.com/soschuetze/Trade-GNN-ChangePoint

- Developed PyTorch Geometric s-GNN model for use in change-point detection of crises related to trade, achieved F1 score of 0.97
- Constructed ensemble method of domain-knowledge and spectral clustering to determine ground-truth economic shock years
- Applied model to region and product subnetworks to understand how crises manifest differently in trade subnetworks

Latent Space Model of Migration-Trade-Terrorism Networks

github.com/soschuetze/ParallelNetworkChange

- Created latent space model to predict co-evolution of trade, migration, and terrorism networks, achieved AUC of 0.91
- Tested temporal network distances such as Frobenius and Minimum Spanning Tree to aggregate migration and trade networks
- Accepted to Networks and Time II Conference hosted by Northeastern University in London

Fine-Tuned Distilbert for BLM Advocacy Classification

github.com/soschuetze/BLM-DistilBERT

- Implemented traditional and deep learning ML models for classifying forms of advocacy contained in 21 million BLM tweets
- Achieved 0.89 F1-score with fine-tuning Tensorflow DistilBERT model, a 25% increase from baseline traditional model
- Analyzed model classifications, discovering shift away from within-the-system and towards disruptive forms of activism

Skills

Languages: Python, SQL, R, Stata, Java, JavaScript, HTML

Tools: TensorFlow, PyTorch, Spark, NumPy, Pandas, Hugging Face Transformers, Scikit-Learn, Git, Matplotlib, NetworkX

Frameworks: Causal inference, hypothesis testing, linear regression, logistic regression, lasso regression, multi-level modeling, neural networks, difference-in-difference, machine learning, deep learning, NLP, Graph ML, data visualization