Sinclaire Schuetze

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

Education

University of Oxford – MSc in Social Data Science, Grade: Distinction (First Class Honours) Aug 2024
Wellesley College – BA in Data Science and Economics, GPA 3.94, Summa Cum Laude, Phi Beta Kappa, Sigma Xi May 2023

Experience

Research Fellow, Stanford RegLab – Stanford, CA Sept 2024 – Present

- Training PyTorch Geometric GraphSAGE GNN to classify taxpayers' risk of tax noncompliance within partnership networks, aiming to improve the IRS's selection process for audits
- Introducing anomaly detection algorithm to understand similarity between all partnerships, augmenting the training data set with information about unaudited partnerships

Data Science Intern, Mercury Insurance Group – Brea, CA Jun 2023 – Aug 2023

- Analyzed XGBoost personal 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 new method of contract testing API endpoints used in creation of data pipelines, making process more efficient for data engineers
- Developed 2 new Rest APIs and updated functionality of existing APIs using Swagger

Software Engineering Intern, JP Morgan Chase & Co. – Wilmington, DE Jun 2021 – Aug 2021

- Automated 6 data pipelines using ETL framework, ingesting and transforming data from various data stores using Spark SQL and JPMC libraries, resulting in a more streamlined experience for business analysts
- Tested pipeline functionality 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 networks, achieved F1 score of 0.97
- Applied model to region and product subnetworks to understand how crises manifest differently in trade

Latent Space Model of Migration-Trade-Terrorism Networks github.com/soschuetze/ParallelNetworkChange

- Implemented latent space model with node2vec to understand if aggregation of trade and migration networks can predict edge creation in terrorism networks, achieved AUC of 0.91
- 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 machine learning 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, leading to finding that advocacy has shifted away from within-the-system and towards disruptive forms

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