

Yang Song

ysong26@ncsu.edu | +1 (217) 419 9133 | <https://github.com/ysong126> | Raleigh, NC 27607

EDUCATION

- **North Carolina State University** Raleigh, NC
◦ Ph.D. in Econometrics/Quantitative Economics (**STEM**) GPA: 3.695 Aug 2018 - Expected May 2024
◦ Research: Applied Econometrics, Machine Learning
- **Boston University** Boston, MA
◦ M.A. in Economics Jan 2015 - May 2016
- **University of Illinois at Urbana-Champaign** Urbana, IL
◦ B.S. in Mathematics and Computer Science Aug 2010 - Dec 2014

EXPERIENCE

- **Capital One Financial Corp.** McLean, VA
Data Scientist Intern June 2023 - Aug 2023
 - Developed an end-to-end deep learning pipeline to predict early risks using data from 4 million credit card customers in 140 segments, outperforming the benchmark model in 70% of the segments with a 40% lower overall prediction error
 - Queried, preprocessed, and validated 24 million rows of credit history data by writing SQL and Python across Snowflake, AWS S3, and JupyterHub
 - Automated a data pipeline by embedding a hyperparameter tuning stage that accomplished model selection using DVC, Python, and Polars
 - Collaborated with data scientists and presented a deck detailing business impacts on a \$9B market and potential enhancement to business stakeholders
- **Shelf Engine Co.** Seattle, WA
Data Scientist Intern June 2022 - Aug 2022
Shelf Engine is a series B startup (~\$60M) that provides inventory management solutions to supermarket chains through machine learning
 - Developed and optimized a deep learning model to forecast market demand for perishable items, improving two key metrics: RMSSE by 15.3% and bias by 10.0%
 - Wrangled 6TB of time series data on Azure Databricks utilizing Python, SQL, and PySpark to evaluate model performance on 15000+ SKU items
 - Presented to the data science lead on transitioning the model from the development to the test phase

SKILLS

- **Programming Languages and libraries:** Python, Keras/TensorFlow, PyTorch, SQL, Shell scripting
- **Misc tools:** PySpark, Polars, Jupyter Notebook, Git, DVC, A/B testing

PROJECTS

- **Forecasting Agricultural Commodity Prices Using a CNN-GRU Neural Network with a Likelihood Loss Function**
 - Designed and built a convolutional recurrent neural network in Python using Keras/TensorFlow to improve forecasting accuracy of agricultural commodity prices, reducing RMSE by 3%
 - Published a module for Keras and Sklearn data preprocessing on Python Package Index (PyPI)
<https://pypi.org/project/lstm-reshaper>
- **Volatility Modeling Using a Hybrid GARCH GBDT Model**
 - Built and fine tuned a gradient boosting decision tree (GBDT) model to forecast market volatility, reducing forecasting error by over 22%
 - Implemented a data pipeline for feature engineering, training, and model evaluation against benchmark models, including stepwise regression, PCA, Lasso, Random Forest, and SVM

PRESENTATIONS

- **American Economic Association Annual Meeting (AEA 2023)**, poster session Jan 2023
◦ <https://www.aeaweb.org/conference/2023/program/paper/kGfAes6K>
- **Midwest Econometrics Group Conference (MEG 2022)**, paper session Oct 2022

TEACHING

- **Independent Graduate Instructor**, Principles of Macroeconomics 2020 - 2022

Coursework and Certificates

- Neural Networks and Deep Learning (Coursera)
- Machine Learning (Coursera)
- Spark and Python for Big Data with PySpark (Udemy)