Nat Fernelius

ferneliusn.github.io

Experience

Staff AI Data Scientist - Tax, Technology, and Transformation July 2023 – Present
Ernst and Young *New York, NY*

- Built AI-assisted data processing pipeline for extracting structured information from unstructured documents such as images of corporate structure charts and legal contracts. This pipeline is currently being integrated into a production application.
- Built embedded VAT tax code prediction model to identify tax booking issues and recommend issue categories of focus for tax compliance teams
- Built graph database powered retrieval augmented generation architecture for use within embedded chatbots and AI agents in a client's central tax application

AI Data Science Intern - Tax, Technology, and Transformation June 2022 – August 2022
Ernst and Young *Dallas, TX*

- Built several production machine learning models for sales and use tax code prediction
- Built real estate price regression model for use within the firm's internal property tax group

Undergraduate Coding Fellowship October 2021– May 2022
The University of Texas at Austin Economics Department *Austin, TX*

- Built ETL pipeline and data processing architecture for a large scale economic dataset whose size necessitated parallel computing methods
- The completed pipeline generated the dataset for a published time series analysis focused economic paper written by a professor at the university on the effects of short term rentals on housing supply.

Education

The University of Texas at Austin Cumulative GPA: 3.96
Bachelor in Economics, Honors *Aug. 2019 – May 2023*
Bachelor in Plan II, Honors, Certificate in Scientific Computation and Data Science *Aug. 2019 – May 2023*

Relevant Coursework Linear Algebra, Econometrics, Parallel Computation

Projects

Predicting Unemployment With Sentiment Measures | *Python, NLP, Time Series Regression, Technical Writing*

- Scraped meeting minutes from the Federal Open Market Committee and generated sentiment measures for each document
- Performed time series analysis on the unemployment rate and found this sentiment measure to be a significant predictor of unemployment at 3 forecasted time horizons
- Published and defended these findings in an undergraduate honors thesis for the economics department

AI-Enhanced Wikipedia Knowledge Graph | *Python, LLMs, Neo4j, PyTorch, Django*

- Created a knowledge graph in Neo4j based on a scraping of select Wikipedia pages
- Finetuned a code instructor model to write Cypher queries against the knowledge graph based on user inputs or agentic feedback from user requests
- Implemented a frontend chatbot experience in Django that allowed users to query the knoweldge graph and receive grounded informed responses to their questions

Technical Skills

Python Libraries: pandas, NumPy, scikit-learn, PyTorch, Django, nltk, Dask
Other Languages: Clojure, Fortran, Bash, C++, SQL, Cypher
Developer Tools: Git, Docker, Linux, Nix
Cloud Services/Databases: Azure (ML Studio, OpenAI, Data Factory), Databricks, Neo4j