

Curriculum Vitae

Chao Peter Yang

Tel: (+1) 773 - 901 - 0080 Email: chaopeter.yang@gmail.com LinkedIn: linkedin.com/in/chao-peter-yang

Personal Website: site-cpyang.github.io/ GitHub: https://github.com/cpyang123

PROFESSIONAL EXPERIENCES

Informa PLC/ Curinos, Senior Technical Success Analyst, Modeling October 2023 – Current

- ◆ Researched and developed industry-level **nonlinear** elasticity models for **Asset-Liability Management (ALM)** to predict acquisition and other portfolio balances for regional banks and credit unions, resulting in improved prediction vs. legacy models in terms of out-of-sample R^2 .
- ◆ Created automated ad-hoc regression notebooks with **PySpark** for creating, testing, and validating models with different configurations, reducing the time to build proof-of-concept models by half.

Informa PLC/ Curinos, Data Science Analyst II April 2022 – October 2023

- ◆ Led ML engineering team to migrate legacy modeling pipeline from using **Cloudera** to **Databricks**, coordinating across teams to schedule testing, promotion, and release plans, leading to more than \$100k in annual savings for platform expenses, and a 30% decrease in data processing time on average. **Acknowledged in company-wide town hall meeting**
- ◆ Tuned **nonlinear hierarchical price elasticity models** en masse for multiple major US banks, each with 10,000+ model segments, resulting in improved fit in terms of both AIC and R^2 with a significantly higher rate of convergence.
- ◆ Set up and automated custom **SQL** procedures to clean, wrangle, and transform client's data feed to be used in the modeling pipeline, eliminating the need for manual model data refreshes.

Informa PLC/ Curinos, Data Science Analyst Aug 2021 – April 2022

- ◆ Converted local, single-threaded, legacy modeling pipeline to use **SparkR** and **Cloudera**, reducing run time for model fitting by up to 30 times.
- ◆ Performed **Exploratory Data Analysis (EDA)** for client banks to tune and reconfigure their models and data segments, leading to better performing **price elasticity models** in terms of MAPE, R^2 , and rate of convergence.
- ◆ Installed and managed more than 10,000 **price elasticity models** per client bank to predict and optimize their deposit portfolio across a wide range of interest rates, with precise **Model Risk Management** documentation.

EDUCATION

Duke University Aug 2024 - May 2026

- ◆ *M.S. in Interdisciplinary Data Science, MIDS Scholarship Recipient* GPA: 3.93 / 4.0
- ◆ Relevant Courses: Natural Language Processing, Statistical Modeling, Data Engineering, Theory of Machine Learning

University of Michigan - Ann Arbor Aug 2018 - May 2021

- ◆ *B.S. in Data Science and Mathematics*
- ◆ Honors: **Highest Honors** in Data Science (**One of Two** in Department), **University Honors** (2019, 2021)
- ◆ Relevant Courses: Linear Algebra, Statistics, Data Structures and Algorithms, Database Management Systems, Numerical Methods, Machine Learning, Mathematical Modeling in Biology, Programming and Data Structures, Introduction to AI

RESEARCH EXPERIENCES

Interpretable Machine Learning Lab, Duke University Aug 2024 – Current
Research Assistant Durham, NC

- ◆ Researched and developed a custom implementation of **DiffPool** for **Heterogeneous GNN** used in musical analysis in **PyTorch**, improving **Cross-Entropy Loss** by more than 60% in validation with additional **hyperparameter tuning**.
- ◆ Developed and implemented **Proximal Policy Optimization** for **Graphical Neural Network** to enable **Reinforcement Learning from Human Feedback**, creating automated personalized musical analysis.
- ◆ Co-advised by Prof. Cynthia Rudin and PhD candidate Stephen Ni-Hahn.

University of Michigan, Honors Student Researcher Sep 2020 – May 2021

- ◆ Developed new music classification methods using Musical Instrument Digital Interface (MIDI) and **LSTM neural networks** resulting in 82% accuracy in music classification, more than 10% improvement over conventional **ML methods**.
- ◆ Improved models using **supervised machine learning methods** like **Support Vector Machines**, **Decision Trees**, **Ensemble Methods**, **K-nearest neighbors** etc. by using expert system for chord identification.
- ◆ Advised by Prof. Edward L. Ionides (Department of Statistics) and Prof. Daniel Forger (Department of Mathematics)

University of Michigan, Student Researcher Jan 2020 – May 2020

- ◆ Led and executed a research project on modern and historical violin strings, employing advanced data analysis to uncover distinct sound quality variations.
- ◆ Utilized **Fast Fourier Transform (FFT)** and **convolution reverb** techniques to simulate acoustic properties of various venues, revealing significant environmental effects on sound perception.
- ◆ Successfully authored and **secured research funding** from Prof. Joseph Gascho, managing financial transactions between vendors and the Stearns Collection at the University of Michigan.
- ◆ Supervised by Prof. Mark E. Newman (Department of Physics)

RELATED PROJECTS

Healthcare Data Analytics Platform - SanAssist Repo Link Fall 2024

- ◆ Built a scalable web app with a team of 4, leveraging a fine-tuned **GPT-2 (LoRA)** to enable healthcare analytics, achieving a perplexity of 3.32 (outperforming Google's MedPaLM), deployed using **Docker** on **AWS ECR** and **App Runner**.
- ◆ Designed an **ETL** pipeline (Databricks, Pandas) with **SQLite** storage and real-time analytics (**Squirrels API**) for healthcare metrics visualization, complete with **CI/CD** using **Github Actions**, and load tested for 10,000 concurrent users.

AI Engineering — Duke ProfMatch Project Link Fall 2024

- ◆ Developed LLM-based professor **recommendation system** for Duke students using **GPT-4o-mini** with state-of-the-art **Graph-based Retrieval Augmented Generation** system, **LightRAG**, allowing for personalized recommendations.
- ◆ Designed and implemented **Streamlit**-based front-end for ProfMatch with full NetworkX visualization of the underlying graphical data and chat-bot-like interface.
- ◆ Scraped profile pages to automatically obtain relevant contact information and profile picture for effortless connections.

Data Engineering — House Price Records Database Fall 2024

- ◆ Developed full-fledged command-line SQL interface allowing for easy CRUD operations with full CI/CD integration via **GitHub Actions**.

Modeling — Muscribe Fall 2023

- ◆ Crafted and deployed **Transformer** and **Convolutional Recurrent Neural Network** models for transcribing audio music to sheet music, achieving near state-of-the-art accuracy with markedly reduced training resources.

Data Engineering — Squirrels API Framework Spring 2023

- ◆ Managed and co-founded the Python package, *Squirrels*, to allow for easy creation of REST APIs with dynamic queries.
- ◆ Developed testing features, allowing for package users to run custom tests to validate query data in the API framework.

Modeling — House Price Prediction Summer 2020

- ◆ Predicted house prices based on real financial data from Kaggle with **supervised machine learning techniques**, resulting in a **boosted-tree** model performing in the top 20 percentile of the class.
- ◆ Performed **feature analysis** and selection on over 10 regression models, resulting in improved performance on the final mode.

Database — Facebook Database Clone Spring 2020

- ◆ Implemented a **relational database** for storing user information of a simulated Facebook-like company using Java for easier query, in addition to a **NoSQL** database to increase its versatility.

Computer Science — Path Optimization Winter 2020

- ◆ Developed a program that solves **Traveling Salesmen Problem** with heuristics and **branch-and-bound** approaches using **dynamic programming** in C++, reducing the solving time by more than 99.95% compared to brute-force methods.

Terminal Command Game — Euchre Fall 2019

- ◆ Implemented a C++ based game of Euchre that supports local PVP and PVC gameplay via terminal command, using **object-orientated programming** with custom classes.

PUBLICATIONS

- ◆ Undergraduate Honors Thesis: *The Classical-Romantic Dichotomy: A Machine Learning Approach*
https://ionides.github.io/students/cpyang_honors_thesis.pdf

AWARDS AND HONORS

- ◆ *Duke University*: MIDS Scholarship (50%)
- ◆ *University of Michigan*: Highest Honors in Data Science for my thesis
- ◆ *University of Michigan*: University Honors (2019, 2021)
- ◆ *Rados Deszö Violin Competition*: Gold Medalist
- ◆ *Central & Eastern European Schools Association (CEESA)*: Silver in Tennis Doubles
- ◆ *Danube Valley Athletic Conference (DVAC)*: Gold in Tennis Doubles

CERTIFICATIONS

- ◆ *DeepLearning.AI*: Neural Networks and Deep Learning
- ◆ *DeepLearning.AI*: Structuring Machine Learning Projects
- ◆ *DeepLearning.AI*: Improving Deep Neural Networks: Hyper-parameter Tuning, Regularization and Optimization
- ◆ *DeepLearning.AI*: Convolutional Neural Network
- ◆ *DeepLearning.AI*: Generative AI with Large Language Models
- ◆ *Google*: Share Data Through the Art of Visualization
- ◆ *DataCamp*: Introduction to Scala

TECHNICAL SKILLS

- ◆ **Languages**: R, Python, SQL, Scala, C++, Javascript, MATLAB

- ◆ **Machine Learning:** SciPy, PySpark, Regressions, Trees, Ensemble Methods, Gradient Descent, Bootstrapping
- ◆ **DNN:** Pytorch, PytorchLightning, MIDI-Toolbox(Deep Learning Package)
- ◆ **Database/Tools:** Spark, Databricks, AWS, SQL Server, PostgreSQL, MongoDB
- ◆ **Visualization:** matplotlib, ggplot, Tableau

VOLUNTEERING & EXTRACURRICULAR ACTIVITIES

- ◆ **Volunteering Experiences:** Greater Chicago Food Depository (2023), Budapest Festival Orchestra (2017), Habitat for Humanity (2015)
- ◆ **Hobbies/Activities:** Model United Nations (ThiMUN 2017, KarMUN 2016), Violin (Kroo Gyorgy Community Orchestra 2015 - 2018, University Orchestra 2018), Tennis (UMich Club Tennis 2019- 2020, High School Varsity 2016-2017)

LANGUAGES

English (Professional), Hungarian (Native), Mandarin Chinese (Native), German (Intermediate) , Japanese (Conversational)

REFERENCES

Prof. Edward L. Ionides, Associate Chair for Undergraduate Studies and Professor

Department of Statistics, University of Michigan
ionides@umich.edu, +1 (734) 615-3332

Prof. Mark E. Newman, Anatol Rapoport Distinguished University Professor of Physics

Department of Physics, University of Michigan
mejn@umich.edu, +1 (734) 764-4437

Ryan Schulz, Client Success & Modeling Director

Modeling Team, Curinos, Inc.
ryan.schulz@curinos.com, +1 (609) 558-0713