

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, *Modeling Analyst II*

October 2023 – Current

- ◆ Researched and developed industry-level nonlinear **Asset-Liability Management (ALM)** models to predict acquisition and other portfolio balances for smaller banks and credit unions, resulting in consistently better 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, significantly 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 with DevSecOps and Application teams to schedule testing, promotion, and release plans, leading to more than \$100k in annual savings for data platform expenses, and an average of 30% increase in pipeline processing time.
- ◆ 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.
- ◆ 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.

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 reconfigured their models and data segments, leading to better performing **price elasticity models** in terms of MAPE, R^2 , and rate of convergence.
- ◆ Set up and automated custom SQL procedures to clean, wrangle, map and transform client's data feed to be used in the modeling pipeline, eliminating the need for manual model data refreshes.

EDUCATION

University of Michigan - Ann Arbor, *BS in Hon. Data Science and Mathematical Sciences*

Aug 2018 - May 2021

- ◆ GPA: 3.65 / 4.0
- ◆ Highest Honors (**One in Two in Department**)
- ◆ 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

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 Data Science) and Prof. Daniel Forger (Department of Mathematics)

University of Michigan, *Student Researcher*

Jan 2021 – May 2021

- ◆ Planned and conducted research to gather acoustic data on modern and historical violin strings.
- ◆ Simulated the acoustic characteristics of different venues and demonstrated the environment's impact on perceived sounds via **FFT** using convolution reverb.
- ◆ Prepared the project proposal and obtained research funding from Prof. Joseph Gascho, as well as coordinated payments between vendors and funding agency through the Stearns Collection at the University of Michigan.
- ◆ Advised by Prof. Mark E. Newman (Department of Physics)

RELATED PROJECTS

Modeling — Muscribe

Fall 2023

- ◆ Developed and implemented both a **Transformer**-based and a **Convolutional Recurrent Neural Network** to transcribe audio music files to sheet music, resulting in predictions comparable to SOTA models, with significantly less training resources.

Data Engineering — Squirrels API Framework

Spring 2023

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

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

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

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

AWARDS AND HONORS

- ◆ *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

LANGUAGES

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

INTERESTS

- ◆ Violin, Fencing, Calligraphy, Badminton, Tennis

REFERENCES

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

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

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