

Yuanhao Geng

200062, East China Normal University, No.3663, North Zhongshan Road, Shanghai, P.R. China

☎ (+86) 136-6193-3341 | ✉ geng29@wisc.edu | 🌐 <https://github.com/GYHHAHA>

Education

East China Normal University

Shanghai, P.R. China

B.S. IN STATISTICS

Sep. 2017 - PRESENT

- Core Courses: Regression analysis, Time series analysis, Multivariate statistical analysis, Mathematical statistics, Bayesian statistics, Nonparametric statistics, C++ programming, Data mining, Stochastic process, Algorithm design and analysis
- Overall GPA: 3.71/4.00
- Major GPA: 3.84/4.00

University of Wisconsin-Madison

Madison, Wisconsin, U.S.

M.S. IN DATA SCIENCE

Sep. 2021 - Jun. 2022 (Expected)

- Offer for 2021 Fall accepted

Skills

Programming Python, C/C++, Java

Accelerated Computing Cython, JAX, numba, RAPIDS

D.S. Relating pandas, scikit-Learn, pytorch, tensorflow

Database MySQL, Redis

Project

Multi-style music arrangement system based on deep learning

P.R. China

PROJECT LEADER, NATIONAL COLLEGE STUDENT INNOVATION AND ENTREPRENEURSHIP TRAINING PROGRAM

Apr. 2020 - Apr. 2021

- **Description:** Music generation is an important subtask for the MIR (Musical Information Retrieval) area. In this program, we develop a neural network for the music generation with multiple styles. Especially, we resolve the problem of Fugue generation, which can be regarded as a promotion for the multi-voice generation task. A paper submission will be made in May for **ISMIR 2021**.
- **Responsibility:** Musical data format cleaning with music21 (a Python package for musical computing tasks), design for the network, encoder and decoder structure, Pytorch implementing

Research on diversification of family business and second generation inheritance

P.R. China

PROJECT PARTICIPANT, SHANGHAI COLLEGE STUDENT INNOVATION AND ENTREPRENEURSHIP TRAINING PROGRAM

Apr. 2019 - Apr. 2020

- **Description:** In this project, with statistical econometrics methods, we make an attempt to figure out the relationship between the changes in the second-generation holding ratio and the diversification of family business.
- **Responsibility:** Data cleaning, statistical modelling and data analysis in Python.

ECNU GO: A campus carpooling platform

P.R. China

PROJECT PARTICIPANT, NATIONAL COLLEGE STUDENT INNOVATION AND ENTREPRENEURSHIP TRAINING PROGRAM

Apr. 2020 - Apr. 2021

- **Description:** In this project, we develop a campus carpooling platform. The original idea comes from the phenomenon that many students in campus are not able to find partners to make a carpooling, especially during the return to school after the vacation.
- **Responsibility:** Yuanhao Geng takes charge of the backend construction with Flask. Database techniques like MySQL and Redis are used throughout the project.

Classification for the effects between protease pair based on graph methods

P.R. China

PROJECT PARTICIPANT, NATIONAL COLLEGE STUDENT INNOVATION AND ENTREPRENEURSHIP TRAINING PROGRAM

March 2021 - PRESENT

- **Description:** Protease plays an important role in the human body, and there will be antagonistic or synergistic effects between different proteases. Given a set of protease gene sequences, extracting the relationship between them is a meaningful task. Our group will implement graph based methods to make inferences on these relationships and the classification problem.
- **Responsibility:** Since this project just starts only for a few days, literature reading is the top priority at this stage.

Internship

Pfizer Pharmaceutical Co., Ltd.

Shanghai, P.R. China

DATA ANALYST

Jul. 2018 - Sep. 2018

- Performing statistical inferences on clinical drug data
- Conducting visual analysis and report production for data from the pharmaceutical department

Institute of Automation, Chinese Academy of Sciences

Beijing, P.R. China

BIOINFORMATICS ALGORITHM RESEARCHER

Jul. 2019 - Aug. 2019

- Improve the existing correlation matching algorithm between micro-RNA and disease targets
- Exploring the application of deep learning methods, especially graph based methods, in bioinformatics

Shanghai Maodian Asset Management Co., Ltd.

Shanghai, P.R. China

RISK CONTROL ALGORITHM RESEARCHER

Jul. 2020 - Sep. 2020

- Using machine learning methods to establish a forecasting mechanism for untrustworthy enterprises.
- Conducting researches on the algorithms for Anti-fraud Transaction Recognition
- Maintaining the database for macroeconomic data.

Awards Selected

2020.4 **Meritorious Winner (<7%)**, Mathematical Contest In Modeling

U.S.

2020.7 **Rank 12/2370**, Tianchi Big Data Competition - Street View House Numbers Recognition

R.P. China

2020.9 **Rank 61/2535**, 2020 iFLYTEK A.I. Developer Competition - Indoor Temperature Prediction

R.P. China

2020.11 **Second Prize**, National University Mathematical Modeling Competition

R.P. China

Open Source Contribution

Contributions for pandas package

- Pandas is a powerful Python package for data computation, manipulation and processing in modern data science. As of Feb 2021, Yuanhao Geng has committed 11 pull requests merged by pandas dev team, including 2 for document optimization, 8 for bug fix and 1 for feature enhancement. Details can be found at <https://github.com/pandas-dev/pandas/pulls?q=is%3Apr+GYHHAHA+is%3Aclosed>.

Joyful Pandas: A complete tutorial for pandas learners

- Joyful Pandas is a tutorial for the learner who desires to master pandas package. It provides comprehensive descriptions for the modern usage for pandas package, including operations of indexing, groupby, reshaping and concating. Besides, it also gives an updated introduction for the pandas data structure, especially after the major version change from 1.0.0.. As of Feb 2021, this tutorial gains 2.2k+ stars on github. More details can be found on the website: <https://github.com/datawhalechina/joyful-pandas>.