

# Kai Li

(626) 841-0005 | [kai.li@stonybrook.edu](mailto:kai.li@stonybrook.edu) | <https://www.linkedin.com/in/garylikai/> | <http://garylikai.github.io/>

## EDUCATION

### Stony Brook University

*Doctor of Philosophy in Applied Mathematics and Statistics (Statistics Track)*

*Master of Science in Applied Mathematics and Statistics (Statistics Track)*

*Advanced Graduate Certificate in Data and Computational Science (and Engineering)*

Stony Brook, NY

*August 2022 - May 2026*

*August 2020 - May 2022*

Tentative GPA: 3.94/4.00

### The Ohio State University

*Bachelor of Science in Mathematics (Theoretical Track)*

*Minor in Computer Information Science (Database Track)*

*Minor in Economics (Theoretical Concentration)*

Columbus, OH

*August 2017 - May 2020*

GPA: 3.672/4.000

## SKILLS

**Programming Languages:** C, C++, Java, L<sup>A</sup>T<sub>E</sub>X, MATLAB, Markdown, Python, SQL, R, SAS, Stata

**Software and Integrated Development Environments (IDEs):** Eclipse, GitHub, Jupyter Notebook, Microsoft Visual Studio, Overleaf, RStudio, TeXworks

**Languages:** English - proficient, Mandarin Chinese - native, Cantonese - native

## ACADEMIC PROJECTS

### Department of Computer Science, Stony Brook University

*Data Science - Understanding Flight Delays*

Stony Brook, NY

*August 2021 - December 2021*

- Retrieved relevant flight arrival performance datasets of interests from the Bureau of Transportation Statistics.
- Preprocessed datasets by subsetting datasets, imputing missing data, merging datasets, and variable encodings.
- Obtained meaningful insights from the datasets by exploring descriptive statistics, testing significant variables, and visualizing data patterns and trends for model building.
- Implemented machine learning models, Ridge regression,  $k$ -nearest neighbors, and neural networks, to model the likelihoods of flight delays and compare their effectiveness.
- Presented the entire research process in a reproducible, tweakable, and well-documented notebook computing environment with an academic report.

### Mathematical Biosciences Institute, The Ohio State University

*Survival Analysis - Epidemic Modelling*

Columbus, OH

*September 2019 - April 2020*

- Developed statistical methods to generate large-population samples from modeling epidemiological processes.
- Analyzed samples segregated into susceptible (S), infected (I), and recovered (R) compartments.
- Generated solutions using ordinary/partial differential equations, survival functions, or cumulative hazard functions.
- Computed the proportion of people susceptible or infected using computer software.
- Interpreted the awareness effect of spreading epidemics under Susceptible-Infected-Recovered (SIR) curves.

**Kaggle Competitions** (<https://www.kaggle.com/>): Microsoft Malware Prediction and Rossmann Store Sales

**PMLi R Package:** Statistical procedures in R to analyze partially matched samples - an experimental design based on independent samples and matched pairs designs.

**Online Bookstore Information Management System and Database:** An online bookstore database system using SQL to support inventory and sales operations.

## EXPERIENCE

### Department of Mathematics, The Ohio State University

*Math Peer Mentor - Four Students in Total*

Columbus, OH

*August 2018 - April 2020*

- Identified possible barriers that students may have on personal, academic, or other problems during the first year to avoid transition issues and adjusting to college life.
- Fostered a sense of community for students and motivated them to utilize campus and community resources.
- Encouraged interpersonal and group interactions among mathematics and actuarial science students to actively participate in volunteering math competitions.

## RELEVANT COURSEWORK

**Statistics:** Exploratory and Categorical Data Analysis, Regression Analysis, Time Series Analysis

**Computer Science:** Data Science, Database Systems, Algorithms, Computer Architecture, Operating Systems

**Economics:** Econometrics, Microeconomic Theory, Macroeconomic Theory