# Kaiyan (Kaylee) Li

kli16@alumni.unc.edu
https://www.linkedin.com/in/kaiyan-li-328268221/


## EDUCATION

**University of North Carolina at Chapel Hill** Aug 2021 – May 2023
*B.S. in Mathematics & B.S. in Statistics and Analytics with Honors (College of Arts and Sciences)*

- Honor Thesis: Robustness of PageRank Centrality on the Undirected Networks
- Graduate with Highest Distinction
- Dean's List for 2021 Fall, 2022 Spring, and 2022 Fall
- Cumulative GPA: 3.89/4.00; Major GPA: 3.92/4.00

**University of California, Irvine** Oct 2020 – Jun 2021
*Major in Mathematics (School of Physical Sciences)*

- Dean's Honor List for 2020 Fall, 2021 Winter, and 2021 Spring
- GPA: 4.00/4.00; Major GPA: 4.00/4.00

## HONOR THESIS

**Robustness of PageRank Centrality on Networks (Honor Thesis)** – Chapel Hill, NC Aug 2022 – May 2023
*Supervisor: Mariana Olvera-Cravioto*

- Modeled random graphs including but not limited to Chung-Lu model and stochastic block model
- Fitted random models to the real-world networks from SNAP to verify the theory and record real-world parameters
- Explored the sensitivity of PageRank centrality to the real-world networks' different graph properties such as degree distribution by comparing real networks and simulated random graphs
- Tested the robustness of PageRank centrality on self-designed mimic real-world undirected networks by simulating random graphs based on fitted real-world parameters
- Concluded that PageRank centrality was very sensitive to the degree distribution, somewhat sensitive to the community structure, and not sensitive to the clustering coefficient
- DOI: https://doi.org/10.17615/93fj-kj91

## WORKING EXPERIENCE

**Cmind Inc** – Boston, MA Jun 2023 – May 2024
*Data Engineer Intern*

- Built and trained machine learning models including Boosting model using weekly updated data to make daily earnings surprise predictions, and significantly minimized errors
- Modified machine learning model by selecting variables with high calculated feature importance to make earnings surprise classifications, and enhanced the accuracy from 65% to 70%
- Generated and visualized model accuracy fluctuation over time using Python and PivotTable
- Wrote the prediction datasets into MongoDB with designed daily update setting
- Designed and changed the data logging mode of database on time-based approach

**Mathematics Student Federation** – Irvine, CA Apr 2021 – Jun 2021
*Tutor*

- Assisted students enrolled in mathematics and statistics courses by answering questions and summarizing core concepts
- Designed mathematical and statistical problems to test students and help them understand materials

## CURRICULUM PROJECT HIGHLIGHTS

**Ice Curling Analysis** – Dept of Statistics and Operations Research, UNC-CH Jan 2023 – Mar 2023
*Supervisor: Mario Giacomazzo*

- Self-determined variables used to analyze and manually collected data through a wide range of games
- Predicted game results based on a self-created weighting system and tested the accuracy of predictions
- Visualized the findings and concluded that the accuracy of self-created weighting system was 92%

**Toxic Comment Classification** – Dept of Statistics and Operations Research, UNC-CH Aug 2022 – Dec 2022
*Supervisor: Yao Li*

- Cleaned the dataset from Kaggle's "Toxic Comment Classification Challenge" through text normalization
- Self-designed and trained three deep learning models Bi-LSTM, TextCNN, and DistilBERT using Pytorch, Keras, and Tensorflow, all reaching around 95% accuracy

- Discovered that DistilBERT outperformed than two other models after comparison with multiple evaluation metrics

**Movies Profit Prediction** – Dept of Statistics and Operations Research, UNC-CH Jan 2022 – May 2022
*Supervisor: Mario Giacomazzo*

- Cleaned the dataset from Kaggle's "The Movies Dataset" through converting JSON objects and mutating variables
- Built models in R to predict movies profit by integrating predictors such as movie genres and in-collection or not
- Concluded that the relative prediction error was about 3% inside the dataset and 7% outside the dataset

## SKILLS

- Algorithm: FFT, SVM, MLP, CNN, RNN, LSTM, BERT, GAN, MDP (POMDP & CMDP)
- Programming: Python, Julia, Dart
- Software: MATLAB, Mathematica, R Studio, LaTeX, SPSS, JMP, Flutter, Excel
- Language: Fluent in English and Mandarin