Yeh Chan Yoo

yehchanyoo@gmail.com | yehchanyoo@berkeley.edu | yehchanyoo.github.io | linkedin.com/in/yehchan-yoo

Education

University of California – Berkeley
Bachelor of Arts in Statistics and Political Economy
Minor in Data Science

Fall 2017 - Fall 2023
Overall GPA: 3.67

Technical Skills

- **Languages/Frameworks:**
  - **Data Science/Statistics:** Python, R
  - **Data Engineering:** SQL (SQLite, Microsoft SQL Server, Oracle SQL)
  - **Data Science Communication:** LaTeX, Markdown
  - **Software Engineering Tools:** Git (version control), Shell
  - **Web Development:** HTML, CSS, Javascript
    - **Web Frameworks:** React.js, Node.js
- **Packages:**
  - **Data Manipulation/Statistical Analysis:**
    - **Python:** NumPy, Pandas, SciPy, Xarray, Statsmodels, ArviZ, Requests, Beautiful Soup, Selenium
    - **R:** dplyr, tidyr, readr, stringr, data.table
  - **Machine Learning:**
    - **Python:** Scikit-learn, PyMC3, PyTorch
    - **R:** carat, e1071, kernlab, mclust, nnet, Mice
  - **Data Visualization:**
    - **Python:** Matplotlib, Seaborn, Plotly
    - **R:** base, ggplot2, igraph

Licenses and Certifications

SQL Developer
Korea Data Agency

Issued June 2022
Expires June 2024

- Demonstrates deep understanding in relational databases, data modeling, and the use of SQL (more specifically, Oracle SQL and Microsoft SQL Server)

Work and Research Experience

UC Berkeley School of Education
Research Assistant

January 2023 - December 2023
Berkeley, CA

- Currently working on neural natural language classification models in Python for automatic short answer grading of student answers to mathematical and statistical questions; achieved 75% test accuracy with the use of a RoBERTa model as of October 29, 2023

- Graded and analyzed written answers to statistics assessment questions from more than 500 California high school and middle school students using Excel and R
- Assisted with the design and writing of the project's statistical assessment questions for California high school and middle school students
- Presented and reported to my project manager, my professors, and other undergraduate research assistants on the topics of machine learning-based automatic short answer grading and the use of GeoGebra for creating interactive educational applets

Republic of Korea Air Force
May 2020 - February 2022
Translator Sergeant
Daegu, South Korea

- Translated hundreds of pages of various government, legal, mechanical, and logistics documents (from Korean to English and from English to Korean) for the central air force depot to facilitate smooth communication on logistical and inventory issues between the Republic of Korea Air Force and various international arms manufacturers such as Boeing, Airbus, and Elbit Systems
- Translated maintenance manuals and training guides for aircraft systems containing extensive aviation and military terminology
- Received recognition from the Republic of Korea Air Force depot leadership for diligence, translation quality, and attention to technical qualities

SK Wyverns
May 2020 - August 2020
Intern
Incheon, South Korea

- Analyzed several thousands of customer records from the baseball team database and from online using Google Analytics, Python, SQL, etc. to identify customer sentiment and their app and website usage patterns, most notably identifying a 67% decrease in session time in the baseball team's fan talk forum after the recent launch of a new version of the fan forum
- Proposed new changes to the website and app user interface based on the analytics data, mainly suggesting the addition of pictures and visualizations and the addition of more links on the homepage to the team's media and forum content for greater accessibility

---

Relevant Teaching Experience

DATA C102: Data, Inference, Decision-Making
August 2023 - December 2023
Tutor
Berkeley, CA

- Writes the rubric and provides individualized feedback for assignment submissions every week
- Independently grades labs, homework, and exams for ~250 students under tight 7-day turnarounds
- Teaches complex data science concepts such as the confusion matrix, random forest, and propensity score weighting to students for at least two office hours every week

DATA C100: Principles and Techniques of Data Science
June 2023 - August 2023
Reader
Berkeley, CA

- Set the rubric and provided individualized feedback for assignment submissions every week

- Independently graded lab and homework assignments for 150 students under tight 7-day turnarounds

Berkeley Math Tournament
November 2022
Berkeley, CA

Exam Proctor and Grader

- Proctored high school participants of the Berkeley Math Tournament – one of the most popular high school math tournaments in the country – to ensure smooth execution and collection of exams
- Graded exam papers from the tournament

DATA C8: The Foundations of Data Science
August 2018 - May 2019
Berkeley, CA

Lab Assistant

- Assisted students with their practice problems, lab problems, and/or project problems in weekly lab sessions

---

Extracurriculars

Open Computing Facility
August 2023 - December 2023
Berkeley, CA

Finance Committee Co-Head

- Manages and plans out spending for the annual $110,000 budget of Open Computing Facility (OCF), an arm of ASUC that runs a server, a computer lab, and a free printing and scanning service; looked over or approved over $20,000 worth of transactions as of November 15, 2023
- Creates complex budgeting plans for the organization's socials, ranging from small hikes to large retreat
- Finalized the operations staff stipend for the 2023-2024 academic year based on a data-driven analysis of financing and economic conditions
- Reported on the current state and the future recommendations for OCF's budget allocation for the 2023-2024 academic year based on data analysis of past transactions and current economic conditions
- Currently working on a financial dashboard for the facility

Open Computing Facility
January 2023 - May 2023
Berkeley, CA

Finance Committee Member

- Analyzed the facility usage metrics over the last ten years to evaluate current facility use by the UC Berkeley community and to estimate the cost needed to run the facility for the 2023-2024 academic year, identifying a 20% decrease in annual printing use compared to pre-COVID years but stubbornly high maintenance costs in computer and server maintenance
- Estimated projected operations staff stipend for the 2023-2024 academic year based on a data-driven analysis of political and economic conditions
- Presented facility usage metrics, stipend calculations, and pain points in computer and server maintenance to apply for funding from student government Associated Students of University of California, successfully resulting in more than $40,000 in additional funding for the 2023-2024 academic year compared to the previous academic year

Student Association for Applied Statistics
Data Consulting Member

September 2019 - December 2019
Berkeley, CA

- Collaborated with 7 other teammates to create several machine learning models in Python for organization and algorithmic understanding of over 50,000 JSON files of complex customer API data from data security startup Trace Data (now acquired by Netskope)
- Led the committee's work in designing a K-means clustering algorithm using Pandas and Scikit-Learn to group the JSON files based on the type of content within the files
- Developed a pipeline that could classify what type of content was contained within the customer data (sales coupons, sports prediction, online orders, etc.) to allow the startup to provide customers with suggestions for keeping their API data secure

Statistics Undergraduate Student Association
Research and Publication Committee Co-Head

January 2019 - May 2019
Berkeley, CA

- Managed and led the committee as a co-head in Spring 2019; guided 15 committee members' research projects by assisting with their proposals, visualizations, and research article drafts while teaching research workshops on topics such as research ethics and data collection
- Organized and spearheaded the club's second (and first science fair-style) research symposium with 100 attendees and 30 projects across 4 data science student organizations to have committee members present their semesterly statistics/data science research projects with one another and with other student organizations

Statistics Undergraduate Student Association
Research and Publication Committee Member

January 2018 - December 2018
Berkeley, CA

- Conducted independent research to determine whether the legalization of euthanasia is connected to the suicide rate in Mexico and the Netherlands in Spring 2018; discovered that the judicial creation of criteria for allowing euthanasia may be connected to the relative decrease in suicide rate in the Netherlands compared to Norway in the 1970s with 10% significance (https://saas.berkeley.edu/rp/suicide)
- Wrote an independent research article on the meaning of probabilities in social sciences in the fall of 2018, discussing how frequentist and Bayesian interpretations of probability can be deployed in social science research (https://saas.berkeley.edu/rp/meaning-of-probabilities-in-social-sciences)

Relevant Undergraduate Coursework

Classes listed in reverse alphabetical order with relevant programming language(s) bolded in parentheses

- UGBA 103: Introduction to Finance
- STAT 156: Causal Inference (**LaTeX, Markdown, R**)
- STAT 155: Game Theory
- STAT 154: Modern Statistical Prediction and Machine Learning (**LaTeX, Markdown, Python, R**)
- STAT 151A: Linear Modelling: Theory and Applications (**Markdown, R**)
- STAT 135: Concepts of Statistics (**Markdown, R**)
- STAT 134: Concepts of Probability
- STAT 133: Concepts in Computing with Data (**Markdown, R, Git, Shell**)
- MATH 110: Linear Algebra
- MATH 104: Introduction to Analysis

- ECON 141: Econometrics (Math Intensive)
- ECON 101B: Economic Theory–Macro
- ECON 101A: Economic Theory–Micro
- DATA C104: Human Contexts and Ethics of Data
- DATA C102: Data, Inference and Decisions (**LaTeX, Markdown, Python**)
- DATA C100: Principles and Techniques of Data Science (**LaTeX, Markdown, Python, SQL**)
- COMPSCI 198: Linux SysAdmin DeCal (Spring 2023) (**Git, Shell**)
- COMPSCI 198: Full Stack Development DeCal (Fall 2023) (**HTML, CSS, Javascript, React.js, Node.js**)
- COMPSCI 188: Introduction to Artificial Intelligence (**Python**)

Volunteering

Disabled Students' Program — February 2023 - May 2023
Disability Notetaker — Berkeley, CA

- Uploaded and shared class notes for STAT 154: Modern Statistical Prediction and Machine Learning with disabled students for academic assistance under the university's guidance

---

Class and Extracurricular Projects

All my class and extracurricular projects can be read on my portfolio website https://yehchanyoo.github.io/.

- "Inference and Prediction on Crude Diabetes Prevalence in U.S. States Based on Vegetable Consumption" by Christina Đặng, Conan Minihan, Tetsuro Escudero, and **Yehchan Yoo**, presented as the final project for DATA 102: Data, Inference and Decisions for Spring 2023
- "Replication and Improvement on 'How do 401(k)s Affect Saving? Evidence from Change in 401(k) Eligibility'" by **Yehchan Yoo** and Xinyi Zi, presented as the final project for STAT 156: Causal Inference for Fall 2022
- "Transit and Housing in California" by Gain Boonvanich, Anita Ding, Yixin Huang, and **Yehchan Yoo**, submitted and presented as the team article for UC Berkeley Datathon for Fall 2022; won second place in Urban Studies Track
- "SAAS x Trace Data" by Amal Bhatnagar, Michael Chau, Pengyuan Chen, Luke Dai, Philip Kabranov, Amanda Ma, Sahil Rao, **Yehchan Yoo**, presented in front of data security startup Trace Data (now sold to Netskope) at the end of Fall 2019
- "Meaning of Probabilities in Social Sciences" by **Yehchan Yoo**, presented at Statistics Undergraduate Student Association Research Symposium at the end of Fall 2018
- "Analyzing Undergraduate Statistics Majors' Preparation in Communication with Non-Statisticians in the University of California, Berkeley" by **Yehchan Yoo**, submitted and presented as the final project for college writing class COLWRIT R4B at the end of Spring 2018
- "Is there a statistical relationship between a region's legalization of euthanasia and that region's suicide rate?" by **Yehchan Yoo**, presented for Statistics Undergraduate Student Association at the end of Spring 2018