

Yuancheng (Kaleo) Cao

San Diego, CA | (858) 255-0381 | yuacao02@gmail.com | [linkedin.com/in/cyc2025](https://www.linkedin.com/in/cyc2025) | github.com/Cao1224 | kaleo-cao.com

SUMMARY

With a foundation in statistical methodologies, excel at end-to-end data processing: from cleaning raw data to deploying machine learning models, driven by curiosity and a passion for actionable insights in data-driven decision-making.

EDUCATION

University of California, San Diego

San Diego, CA

- **Bachelor of Science in Data Science**, GPA: 3.65/4.00 Expected March 2025
- **Bachelor of Arts in Interdisciplinary Computing and Arts**
- **Selected Coursework:** Data Structures and Algorithms | Statistical Methods | Probabilistic Reasoning and Learning | Scalable Data Systems | Web Mining and Recommender Systems

CodePath, certificate in Web Development, iOS Development, Android Development

June 2022 – November 2023

Cornell University, online certificate in Machine Learning Foundations

August 2024

SKILLS

- **Programming Languages:** Python, JavaScript, R, HTML/CSS, SQL, C/C++, Java, Swift, Kotlin
- **Libraries:** React, Ray, scikit-learn, Pandas, TensorFlow, Keras, NumPy, Matplotlib, d3.js
- **Language:** Fluent in English and Mandarin

PROFESSIONAL EXPERIENCES

Software Engineer Intern

Remote

CodeDay

June 2023 – August 2023

- Optimized file organization in AWS S3 using TypeScript by implementing intuitive, user-friendly tagging system
- Implemented a custom API for AWS S3 objects, providing seamless CRUD operations for file metadata, resulting in a 40% improvement in data retrieval speed and improving overall system efficiency by 25%
- Collaborated closely with front-end engineer to integrate AWS S3 for a powerful file management solution
- Conducted testing to ensure compatibility across various cloud providers and improve system reliability

PROJECTS

Recommendation Strategy of Kuaishou | Python, Scikit-learn, Pandas

October 2023 – December 2023

- Analyzed digital platform user behavior to predict retention, enhancing long-term engagement and satisfaction
- Standardized engagement metrics to create composite scores that better reflect real user interactions
- Developed a predictive model using Gradient Boosting, significantly achieved 87.4% accuracy in user retention

Global Soccer Visualization | JavaScript, d3.js

June 2023

- Built advanced visualizations to uncover insights from an international football dataset of 44,000+ matches
- Designed an interactive guidance button for charts, significantly enhancing user comprehension and engagement
- Highlighted and annotated key data points with comments, clarifying chart's pivotal information for viewers

TastyData | Python, Scikit-learn, Pandas, matplotlib

February 2023 - March 2023

- Analyzed a recipes dataset to assess changes in protein content over time, identifying 12% increase over 10 years
- Investigated missing data patterns and found a 0.6 correlation between average ratings and factors like preparation time and ingredient variety
- Used GridSearchCV to optimize variables, reducing RMSE by 15% between training and testing sets

Open Energy Dashboard

October 2022 – December 2022

- Improved the interface by fixing the button order and visually highlighting the current page, resulting in a smoother user experience and reducing user navigation errors by 15%
- Implemented JSDoc to enhance code comprehension and facilitate effective communication among developers