

HAOTIAN CHEN

Data Analyst

htc0511@berkeley.edu | 510-926-8047 | Albany, CA | [linkedin.com/haotian-chen](https://www.linkedin.com/haotian-chen)

EDUCATION

University of California, Berkeley

Berkeley, CA

Master of Analytics

Expected May 2025

Coursework: Analysis and Design of Databases (SQL), Optimization Analytics (Python), Risk Modeling & Simulation (Python), Machine Learning and Data Analytics (Python)

Boston University

Boston, MA

Bachelor of Arts in Mathematics and Computer Science

May 2023

SKILLS & INTERESTS

- **Languages:** Mandarin, English.
- **Computer Skills:** Python (Scikit-Learn, Pandas, Matplotlib, Gurobi), R (Tidyverse, Dplyr), Tableau, SQL, Excel (Power Query, Pivot Table), Go, Javascript, React.
- **Technical Skills:** Data Cleaning, Data Visualization, Statistical Analysis, Database Management, Machine Learning, Natural Language Processing, Algorithmic Analysis
- **Interests:** Graphic design (Rhino, AutoCAD, 3DSMax, Photoshop), Chinese Brush Painting, Billiards

INTERNSHIP EXPERIENCE

Alphabet

Mountain View, CA, US (Remote)

Educational Product R&D Intern

Dec. 2023 – Feb. 2024

- Performed market surveys, competitive analysis, and user analysis for 8 popular language-learning softwares and formulate product development features and suggestions.
- Planned the design of UI, data flow, and use cases through Xmind for a mobile Chinese language learning application.
- Constructed UI templates graphically using Pixso for 5-tab and 10-page implementations.
- Implemented the front and back-end signup and login function using React in HBuilder and supported account data storage using MySQL and used advanced queries to communicate between software and database.

Microsoft Research Asia (MSR Asia)

Beijing, China

Data Analyst Associate Intern

Nov. – Dec. 2023

- Processed a department dataset of over 200,000 records of system metric behaviors of 191 Android malware families using Python for models to predict malware families.
- Built classification tree models to predict malware families and achieved over 65% test data accuracy and optimized tree structure parameters using cross-validation through GridSearchCV and gained a 4% increase in test prediction accuracy.
- Deployed from scratch a learning neural network model with 1 hidden layer using forward and backward propagation and achieved test prediction accuracy of 74%.
- Delivered 4 progress reports and 1 final presentation on model performances and significance.

Ningbo Anjie Electronic Technology Co., Ltd.

Ningbo, Zhejiang, China

Data Analyst Intern, Research and Development Department

May – Oct. 2023

- Recorded and labeled product usage data and medical records from over 3,000 customers in R for optimizing the classification models of a commercial sleep-monitoring device in mattresses to alarm abnormal sleep behaviors.
- Performed feature selection and identified 3 trivial variables recorded: body temperature, bodily movement, and height and applied support vector machine model in Python to classify abnormal sleep behaviors given sets of users' physiological and vital data collected by the device; achieved an average classification score of 71%.
- Achieved 89% high-customer-satisfaction responses from experimental surveys of the demo and released the product for sale and signed production contracts with 3 industries.

PROJECTS

Machine Learning Career Roadmap Website

Berkeley, CA

Group Project, Project Manager

Jan. 2025 – May 2025

- Fine-tuned LLMs for extracting skills from over 15000 scraped job description data and over 8000 resume text data.
- Computed word-embeddings and cosine-similarities among 8000 job titles and 8500 skills to first cluster job titles and then construct hierarchy graphs to reflect skill relationships such as prerequisite and association.
- Formulated an ILP optimization process for determining the most suitable courses and learning schedule for users.
- Established 11 frontend UI pages using next.js and connected Fast API backend server with MySQL database.

Berkeley Mobile Food Truck Database Design

Berkeley, CA

Group Project, Project Coordinator

Oct. – Nov. 2024

- Designed the database schema of a fictional mobile food truck startup business using EER diagrams incorporating 13 entities.
- Modeled the schema in MySQL and devised advanced SQL queries to identify operational risks and efficiency.
- Formulated the truck route and scheduling optimization as a linear program and solved it using Gurobi to optimize service.