

Du Xiang

(+1) 858-666-5552 | xd00099@berkeley.edu | <https://du-xiang.com>

EDUCATION

University of California, Berkeley

08/2022 – 05/2023

Master of Engineering in EECS, Data Science and System track

Honors: Received **20,000 dollars** Fung Institute Leadership Scholarship

Courses: ML Design on Multimedia, Parallel Computing, Deep Reinforcement Learning, Computer Architecture

University of California, San Diego

09/2018 – 06/2022

Bachelor of Science in Data Science, GPA: **3.99/4.0**

Honors: **Summa Cum Laude**, Data Science Scholar's Award (By Nomination)

Courses: Data Science and Systems, Data Structure & Algorithm, Machine/Deep Learning, Optimization, Web, Database

SKILLS

Programming Languages: Java, Python, C++, SQL, JavaScript, HTML, CSS, Bash

Skills & Tools: Big data analytics, Spark, Hadoop, Dask; Full-stack web development, Node, Redux, React, HTML/CSS, JavaScript; Machine Learning, Pandas, PyTorch, Scikit-learn; Model development/deploy, Kubernetes, Docker, Heroku; App Development, React Native, AWS. Specialized in text mining and NLP. Proficient working as a team with Git.

EXPERIENCE

UC Berkeley Industry Capstone | Zendar

03/2020 – 06/2022

- Applying state-of-the-art **deep learning** and **computer vision** models to integrate Zendar's high resolution 3D radar data with camera-based perception methods to solve novel **autonomous driving** problems.

Machine Learning and NLP Intern | San Diego Supercomputer Center

02/2021 – 08/2022

- Collaborated in the machine learning team on developing the synthetic biology knowledge system (NSF funded)
- Applied state-of-the-art language model **BioBERT** to extract 9000+ relations with **92+**% accuracy ([co-publication](#))
- Experimented with **Topic Modeling** trained on 1000+ academic articles: achieved high topic interpretation score rated by domain experts. Constructed **graph database** using **Neo4j** that unions 2 unrelated cluster topics
- Pipelined Text Parsing code that converts 2000+ publication PDFs into well-parsed JSON files
- Containerized different tasks using **Docker** which brought **20x+** speed up for deploying training tasks

Instructional Assistant | Halicioğlu Data Science Institute

03/2020 – 06/2022

- Department honored IA for *Intro Machine Learning; Data Science Principles; Practice of Data Science & ML*
- Held office hours, provided technical support, worked on assignment **release engineering & code debugging**

Data Science Researcher | CNS Laboratory of Memory and Brain

01/2020 – 06/2021

- Served as a data science researcher who works on **3D brain images** analysis and **task automation**
- Data cleaning/image processing: filled in 20% of missing data using **statistical sampling**, **automating** image edits
- MRI image analysis: applied **decision tree** to filter top 5 brain features, and perform regression with amnesia rate
- Wrote an **automated pipeline** that boosts WMH volume segmentation efficiency **10+x** w/ **Bash and Python**

PROJECTS

Social Wall (Full Stack (MERN) Web Application | MongoDB, Express.js, React.js, Node.js, Redux.js | [website](#))

- Built a **cross-platform** social app that supports online sharing, commenting, searching, and recommendation.
- Implemented a modern **authentication** platform that supports native signup, login, and Google **OAuth2.0**.
- Troubleshooted bugs from feedbacks both in client ends and server ends; deployment using **Heroku** and **Netlify**

Faculty Information Allocation System (Topic Modeling / NLP | Python, Flask, Docker, Kubernetes | [website](#))

- Applied unsupervised learning (LDA) to facilitate industry allocation among 70+ professors; presented to the board
- Lead a backend team of 3 to build pipelines including **Data ETL**, **Preprocess**, **Modeling**, and **Web Deployment**
- Deployed the product as a web app that integrates the fine-tuned **model**, **search tools**, and **Sankey visualization**

Image Captioning Tool (Deep Learning Application | PyTorch, Python, Flask, Heroku | [website](#))

- Built an encoder-decoder deep learning framework to train on 100K Coco images and generate caption sentences
- Experimented with VGGNet, ResNet (encoder) + LSTM, RNN (decoder); final model achieved **0.69+ BLEU**
- Designed interface and deployed the light-weight model as a playable web-app using Flask, Heroku

Pocket Health App (Mobile App Development | React Native, AWS, GraphQL, JavaScript)

- Student-initiated project that aims to help the people in need of free healthcare (a team with Med and CS students)
- Implemented the reusable components using **React-Native** and **JavaScript** to cut down 30% code length
- Built user authentication system w/ **AWS Cognito** and redesigned the app status logic with 20% less response time