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 Electrical Engineering and Computer Science, Data Science & Systems track
Honors: Received 20,000 dollars Fung Institute Leadership Scholarship
Courses: ML Design on Multimedia, UI/UX, Deep Reinforcement Learning, Parallel Computing, Computer Vision

University of California, San Diego   **09/2018 – 06/2022**
Bachelor of Science in Data Science, GPA: **3.99/4.0**
Honors: Top 1%, Summa Cum Laude, Data Science Scholar's Award, Outstanding TA
Courses: Data Science and Systems, Data Structure & Algorithm, ML/DL/NLP, Web, Relational Database, Optimization

EXPERIENCE

Research Assistant | UC Berkeley & Zendar   **09/2022 – continue**
- Researching with startup team on radar data and methodology to fuse information from radar and camera
- Designing and implementing **deep learning** fusion architectures to integrate 3D radar point cloud with camera-based perception sensor for object detection and solving computer vision and **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)
- Leveraging **LDA Topic Models** trained on 1000+ academic articles with applied **NER** terms: achieved high topic score rated by domain experts. Constructed **graph database** using **Neo4j** that unions 2 unrelated data clusters
- Manually scraped 1000+ articles from PubMed using **Beautiful Soup** as additional model benchmark data
- Containerized various tasks w/ **Docker** which brought **20x+** speed up for deploying training tasks w/ **Kubernetes**

Data Pipeline, Python Developer Intern | CNS Laboratory of Memory and Brain   **01/2020 – 06/2021**
- Served as an agile python developer who works on **3D brain images pipelines**, **analysis**, and **task automation**
- MRI image analysis: applied random forest to filter top 5 brain features, and perform regression with amnesia rate
- Engineered a **data cleaning** and **image processing pipeline** to create automated data visualizations from MRI data
- Wrote a pipeline on **Linux** server that boosts WMH volume segmentation efficiency **10+x** w/ **Bash and Python**

PROJECTS

SocialWall (Independent | GitHub: https://github.com/xd00099/SocialWall | Website: https://social-wall.netlify.app/)
Created a Full-Stack SocalWall web application for friends to share their daily activities.
- Built a full-stack **cross-platform** social website that supports online sharing and deployed it onto Netlify (PaaS).
- Designed and implemented the Instagram-like frontend view of home page and main page using **React.js**
- Implemented **authentication** system that utilized **MongoDB** to store and hash user credentials; added Google **OAuth2.0** support for login; applied middleware to handle permissions; managed the app states with **Redux.js**.
- Implemented the RESTful APIs with **Express.js** and **Node.js** to handle HTTP requests and responses
- Devised database schema for storing user accounts and posts using MongoDB, deployed it on Heroku for stability

Faculty Allocation System (Group, Backend Leader | Website: https://xd00099.github.io/hdsi_faculty_tool/)
Constructed a software tool from scratch that enables faster faculty allocation for industry needs.
- Manually scraped publication data and google scholar labels and preprocessed for natural language processing
- Trained LDA models that automatically and successfully labels and categories each faculty based on publications.
- Lead a backend team of 3 to build pipelines including **Data ETL**, **Preprocess**, **Modeling**, and **Web Deployment**

Image Captioning Tool (Group Leader | GitHub: https://github.com/xd00099/Image-Captioning-LSTM)
Built a deep learning application to automatically generate caption sentences based on provided images.
- Engineered an encoder-decoder deep learning framework using **PyTorch** and trained on 100K Coco image dataset
- Experimented with VGGNet, ResNet as encoder, LSTM, RNN as decoder; final model achieved **0.69+ BLEU**
- Utilized Random Search for hyperparameter tuning and model selection, highly improved optimization efficiency
- Designed interface and deployed the light-weight model as a playable web-app using Flask, Heroku

Pocket Health Mobile App (Student Organization)
- Implemented the reusable components using **React-Native** and **JavaScript** to cut down 30% code length
- Built user authentication system w/ **AWS Cognito** and redesigned logics w/ **React hooks** to reduce 20% loading time

SKILLS

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

Skills & Tools: Machine Learning, PyTorch, Scikit-learn, Kubernetes, Docker; Big data analytics, Spark, Hadoop, Dask; Web development, Node.js, React.js, HTML/CSS, JavaScript; Text mining and NLP. Proficient working with Git.