Du Xiang

+1 (858) 666-5552 | xd00099@berkeley.edu | Website: http://du-xiang.com | LinkedIn: https://linkedin.com/in/du-xiang/

EDUCATION

University of California, Berkeley, Electrical Engineering and Computer Science — **2022 - 2023**
Master of Engineering in EECS, Data Science & Systems Track
Honors: Fung Institute Excellence Scholarship (**20,000 dollars**)

University of California, San Diego, Halıcıoğlu Data Science Institute — **2018 - 2022**
Bachelor of Science in Data Science | GPA: **3.99/4.0**
Honors: **Summa Cum Laude**, **Data Science Scholar's Award** (Faculty Nomination)

TECHNICAL STRENGTH

Programming Languages: Python, Java, SQL, Bash, Javascript, HTML/CSS, C, R, MATLAB
Software & Tools: Kubernetes, Docker, Pandas, React Native, Scikit-learn, Jupyter, PostgreSQL, Neo4j, TigerGraph, Apache Spark, Dask, AWS, Matplotlib, Seaborn, PyTorch, Keras, Git, Tableau, Matlab, Anaconda, Vim
Text mining & NLP: Web crawling with Scrapy/Beautifulsoup, word cleaning, Tfidf, Topic Modeling, Relation Extraction, Named Entity Recognition, XML Parsing, Bert, NLTK, Language Models, Mallet

COURSES

Theories and Applications of Data Science; Advanced Algorithms and Data Structures; Recommender Systems; Convex Optimization; Data Visualization; Probability and Statistics; System for Scalable Analytics; Machine Learning; Deep Reinforcement Learning; Parallel Computing (in 2023); User Interface Design and Development (in 2023);

EXPERIENCE

UC Berkeley Capstone Project Partner | Simplr, Asurion — **09/2022 – continue**
- Generating customer service response template using state-of-the-art language models
- Creating a generative conversational agent that would solve certain customer inquiries coherently and accurately

Machine Learning and Natural Language Intern | San Diego Supercomputer Center — **02/2021 – 08/2022**
- Applied **state-of-the-art language models** (BioBERT) to **extract relations** between chemicals and genes
- **Topic Modeling** trained on 1000+ academic articles: achieved high topic interpretation score rated by domain experts. Results were used to find similar articles and build **author-article network**.
- Organized Text Parsing code (**XML parsing**) and **containerized** different tasks using **Docker**.

Instructional Assistant | Halıcıoğlu Data Science Institute, UC San Diego — **03/2020 – 06/2022**
- IA for *Theories for Intro Machine Learning*; *Data Science Principles*; *Practice of Data Science & ML Pipeline*
- Help with structuring classes, holding **office hours**, and **debugging** in programming assignments on Gradescope

Undergraduate Researcher | CNS Laboratory of Memory and Brain — **01/2020 – 06/2021**
- **Data cleaning/image processing**: worked on fill in missing values using statistical method, automating image edits
- Brain **data visualization** and **statistical analysis**: visualized using **matplotlib**, experimented **correlations** with **PCA**
- Wrote an **automated data pipeline** for WMH volume analysis for 50+ subjects using **Bash and Python**

Network Optimizer Intern | China Telecom, Suzhou, China — **06/2019 – 08/2019**
- Worked with wireless network optimization including **coverage, band frequencies, locations**, etc.
- Analyzed and cleaned user data using **Pandas and Python** to assist creating network solutions

PUBLICATIONS

McInnes, Bridget, **"Discovering Content through Text Mining for a Synthetic Biology Knowledge System"** ACS Synthetic Biology 2021 (ACS Synth. Biol. 2022, 11, 6, 2043–2054)
- Contributor of the **Topic Modeling** & **Relation Extraction** parts

PROJECTS

Faculty Information Retrieval System (Topic Modeling / NLP | Python, Flask, Docker, Kubernetes | website)
- Used unsupervised learning (LDA) to categorize faculty members for industry needs; presented to the board
- Implemented the backend structure including Data ETL, Preprocess, Modeling, and Web Deployment
- Built a dashboard web app that integrates the fine-tuned model, search tools, and Sankey visualization

Image Captioning Tool (Deep Learning / Web App | PyTorch, Python, Flask, Heroku)
- Built an encoder-decoder deep learning framework to identify images and generate caption sentences.
- Optimized ResNet + RNN/LTSM with 30+ sets of hyperparameter; deployed as a web-app using Flask, Heroku

Pocket Health App (IOS / Android App Development | React Native, AWS Amplify, GraphQL, JavaScript)
- Initiated a project that aims to help the people in need of free healthcare (a team with Med and CS students)
- Designed and implemented the reusable components for the front end using React-Native and JavaScript
- Built user authentication system using AWS and designed the backend database schema in DynamoDB

Interactive Author-Articles-Institutions Graph Network (Research / NLP | Neo4j, Aura Cloud, SQL)
- Constructed a graph relational database to show graphical relationships for 2000+ publication data
- Used text mining to bring together PubMed data and ACS data to highlighted novel topic clusters