Zhihao **Du**

zhihao617@berkeley.edu | +1 (510) 833-4417 | GitHub | Personal Website

Education

M.Anlytx (Analytics) – **University of California, Berkeley** 08/2023 - 05/2024
GPA: in progress Berkeley, CA

B.A. Computer Science, Statistics – **University of California, Berkeley** 08/2019 - 05/2023
GPA: 3.8/4.0 Berkeley, CA

Skills

Languages – Python, SQL, Java, C, R, C#, HTML+CSS.

Coursework – Natural Language Processing, Machine/Deep Learning, Statistical Learning, Service Operations Design, Database Systems, Reproducible Data Science, General Linear Models, Data Structures, Machine Structures, Linear Algebra, Probability & Statistics.

Tools/Frameworks – Python Frameworks: Numpy/Pandas, Matplotlib, Scikitlearn, Pytorch/Jax, Flask (since 2023). MS Azure, MySQL Database, MongoDB, .NET, DBeaver, Git, Unix System, Microsoft Office.

Professional Experience

ETL Engineer Intern – **DataCVG Co Ltd** 05/2021 - 08/2021
Relational database design and extract-transform-load pipelines with SQL Shanghai, China

- Merged and resolved client FosunPharma's two conflicting relational database into one new consolidated database by designing and implementing extract-transform-load (ETL) pipelines with DBeaver and SQL;
- Submitted and peer-revised 100+ complex SQL query scripts for data merger pipelines constructed using DBeaver;
- Debugged failures and overcame data confidentiality through extensive communication with client representatives;
- Increased client's database system throughput by 80% benefiting downstream business intelligence tasks.

Academic/Research Experience

Research Assistance – **Project AEI** 01/2022 - present
Real-time audio emotion classification Berkeley, CA

- Advised by Prof. Dacher Keltner, experimented, trained, tested, and finetuned a parallel CNN Transformer neural network using pytorch and MS Azure to classify emotions from streaming real-life audio speech data;
- Spearheaded training data preprocessing with robust data augmentation techniques including Gaussian white noise, simulated room impulse response, and randomly sampled background noise.
- Boosted testing accuracy from 68% to 71% and ensured consistent evaluation performace on real-time audio inputs;
- Programmed and installed real-time audio streaming and continuous model evaluation on a Raspberry Pi 4 device by downsizing audio sample rate and performing batched evaluation to accomodate for the resource-limited platform.

Tutor – **University of California, Berkeley** 01/2023 - present
CS182/282A: Deep Neural Networks Berkeley, CA

- Revised, improved and consolidated interactive demos on BERT with motivating surrogate tasks;
- Developed and peer-reviewed new content material on CNN concepts and applications;
- Led, facilitated and supported students on weekly discussion sections and homework parties

Technical Projects

Howamidoing Full stack web application development with Flask 01/2023 - present

- Designed and developed college level course grade tracker and class standing estimator using the Flask framework, HTML, CSS, and JavaScript;
- Implemented and optimized JSONizable user data objects and stored in NoSQL database with connection to MongoDB, locally and through MongoDB Atlas

Zilean Python package for data mining and engineering pipelines 05/2022 - 08/2022

- Advised by Prof. Fernando Pérez, developed python package "zilean" that bridges the Riot Games API with traditional python data science APIs (scikitlearn, pandas) to produce data pipelines for multidimensional data ready for downstream ML or DL tasks;
- Programmed, tested and refined data mining/engineering algorithms for large semi-structured with rate limiting API request algorithms;
- Promoted and published as open source project with immediate collaborators after established CI/CD pipelines using Github Actions and Readthedocs documentation